Regional Summer Climate Variability and Change over North America in the 20th Century

by

Catherine Anna Nikiel

B.S. Civil Engineering, Rice University (2016)

Submitted to the Department of Civil and Environmental Engineering in partial fulfillment of the requirements for the degree of

Master of Science in Civil and Environmental Engineering

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

February 2018

Signature redacted

Author . . . . . .

Department of Civil and Environmental Engineering

January 19, 2018

Signature redacted

Certified by . . . . . .

Elfatih A.B. Eltahir

Breene M. Kerr Professor of Hydrology and Climate

Thesis Supervisor

Signature redacted

Accepted by . . . . . .

Jesse Kroll

Professor of Civil and Environmental Engineering

Chair, Graduate Program Committee

Regional Summer Climate Variability and Change over North America in the 20th Century

by

Catherine Anna Nikiel

Submitted to the Department of Civil and Environmental Engineering
on January 19, 2018, in partial fulfillment of the
requirements for the degree of
Master of Science in Civil and Environmental Engineering

Abstract

The climate of North America reflects the natural variability of global atmospheric circulation, and is impacted by two primary anthropogenic forcings: increasing concentrations of greenhouse gases (GHG) and Land Use/Land Cover Change (LULC). The objective of this thesis is to quantify the relative impacts that these two anthropogenic forcings have had on summer climate over North America in the course of the 20th century. Simulations are conducted spanning the full 20th century over the United States to the east of the Rocky Mountains, and analysis of model simulations and observations is carried out with a focus on the Midwestern United States. This region has exhibited significant changes in climate from the early to late 20th century, and has been impacted not only by increasing GHG concentrations but by rapid and extensive agricultural LULC changes. These land use changes can be broken into three components: expansion of agriculture, expansion of irrigated cropland, and agricultural intensification in the form of linear productivity gains for corn and soybean.

The simulations in this thesis use a modified version of the MIT Regional Climate Model (MRCM). New input data sets are used in order to represent contemporaneous and evolving lateral and surface boundary conditions. Studies in the past have shown a link between LULC changes and regional/local climate change, and in particular cropland-influenced climatic change in regions with high agricultural intensity. This thesis adopts an approach involving the application of long-term simulations of historic LULC evolution in conjunction with appropriate scenarios of GHG concentrations.

Two sets of simulations are run using either ERA20C or CERA20C 20th century reanalysis data from the European Centre for Medium-Range Weather Forecasting. The simulations are conducted from 1900 or 1901 to 2005 based on data availability, at 30 km resolution on 18 vertical pressure levels. Analysis of changes from 1920-1949 to 1970-1999 show that in most cases, the regional climate impacts from vegetation development in the study area over this period outweigh the impacts from greenhouse gases, but in all variables except temperature, the combined impacts of LULC and GHG alone are not adequate to reproduce the observed climate changes. The analysis of background conditions points to the potentially strong impact of internal variability, and the strong influence that the assumed boundary conditions based on reanalysis data can impose on model results.

Thesis Supervisor: Elfatih A.B. Eltahir
Title: Breene M. Kerr Professor of Hydrology and Climate

Acknowledgments

I owe my ability to complete this work and submit this thesis to many people. First, thank you to my advisor Dr. Elfatih Eltahir who has been attentive to my work and who has helped me to grow as a researcher in the time that I have been here. I look forward to the things that I will learn under your continued mentorship. Thank you also to my colleagues and friends in the Eltahir research group and Parsons Laboratory for welcoming me so warmly and for sharing their knowledge, and especially Mohamed Siam, Suchul Kang, Ross Alter, and Hamed Ibrahim for their patient and generous instruction and advice regarding the inner workings of MRCM.

I would also like to thank Karl Hennermann at the European Centre for Medium-Range Weather Forecasting (ECMWF) for helping me to obtain the data I needed to run my simulations, and Colleen Mikovitz at the National Aeronautics and Space Administration (NASA) for clarifying the NASA Surface Radiation Budget dataset.

Thank you to my friends - who have been extremely supportive and willing to listen, and who have always been there to make me laugh and take a breath. Your support and friendship is so important to me. Finally, I would like to thank my parents, without whom I would not be here. I am blessed to have your support and love, and I am so glad that I can share this with you.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Climate change is a multifaceted issue that affects a complex system. It is anthropogenically driven on a global scale — the chemical composition of the atmosphere is altered by emissions of greenhouse gases and other activities — and it impacts a system with inherent natural variability in response to external forcings. As we continue to deepen our understanding of the Earth's climate, we discover new ways that internal processes interact and influence one another, and further ways in which the climate responds to perturbations on a variety of scales.

Climate change does not just have a global scale cause-and-effect. Humans impose further changes to the environment at local and regional scales through land use and land cover conversion. Aside from direct impacts such as desertification and coastal flooding due to wetland loss, these conversions alter the carbon, water, and energy balances of regional biomes, and are now being studied with models that can more accurately represent fine-scale variations. Land use changes have only recently begun to be robustly included in climate modeling ensembles and aggregate reports on the current scientific state of climate research, such as those produced by the Intergovernmental Panel on Climate Change (IPCC) (IPCC, 2014). The exclusion of these factors meant that previously proposed mitigation and adaptation solutions may have underestimated the locations at greatest risk and the magnitude of potential climate change. The true effects and mitigations for climate change will happen on regional and local scales and must be studied at these scales to understand the unique conditions and development that will occur in countries and states.

One of the most important types of human land cover change is the spread of agriculture. Agricultural land — arable land, and land occupied by permanent crops and pastures — occupies a substantial portion of the Earth's land area. In 2014, nearly 49 million square kilometers of agricultural land covered 37.5% of land area on Earth (Figure 1-1 & 1-2) (World Bank, 2017). The distribution of these areas varies by continent but there are substantial

agricultural areas located on every inhabited continent (Figure 1-3) (Roser & Ritchie, 2018). The percentage of arable land area by country in 2014 is shown in Figure B-1 in Appendix B. The cultivated areas on Earth have expanded rapidly over time, especially in the last several centuries, keeping pace with a growing population (Figures B-2 & B-3) (Roser & Ortiz-Ospina, 2018; Roser & Ritchie, 2018). Despite this, the actual percentage of agricultural (Figure B-4) and arable land (Figure B-5) use per person has decreased over the last several decades as productivity and yields have increased (Roser & Ritchie, 2018).

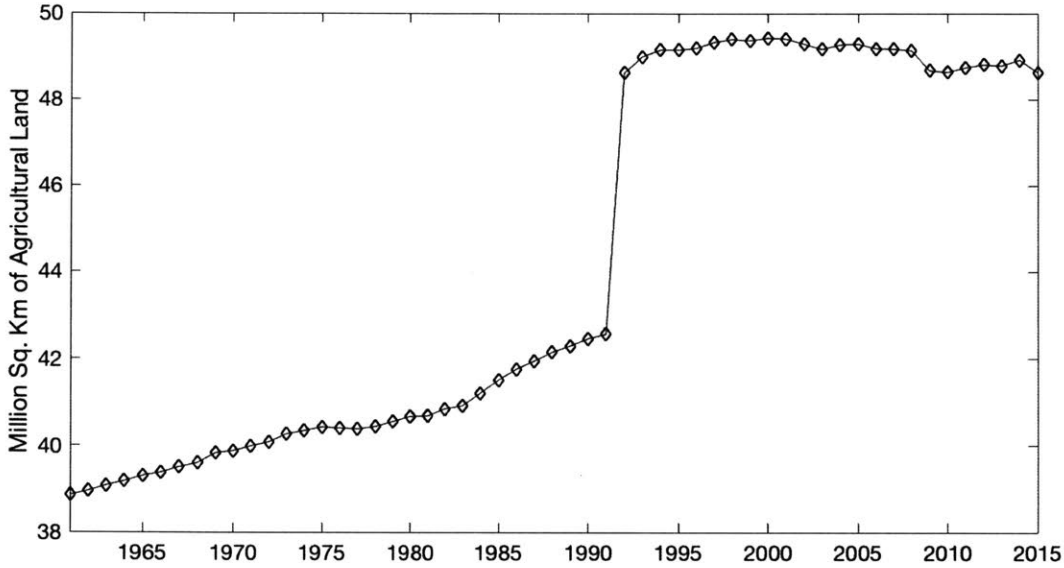


Figure 1-1: Kilometers of agricultural land shown from 1961-2014 (Figure Source: World Bank, World Development Indicators (World Bank, 2017). Data Source: Food and Agriculture Organization)

As reported in the most recent World Population Prospects report from the United Nations Department of Economic and Social Affairs, the world population is expected to rise from the 2017 value of 7.6 billion to 11.2 billion by 2100 (UN, 2017). As these trends continue, the world's agricultural regions will either need to expand or become more productive. However, cropland area has been decreasing in recent decades in population dense areas (Figure B-2) (Roser & Ritchie, 2018). Similarly, productivity trends show potential danger, as there is evidence that yields have been stagnant or declining in nearly a third of the world's key crop areas (those with maize, soybean, rice, and wheat cultivation) over the past half century (Ray et al., 2012). There are still substantial areas where yields are improving, but there are questions about whether this will continue, or if these areas will be able to contribute enough to satisfy needed growth (Ray et al., 2013).

While overall yield trends are important, interannual variability of yields introduces a

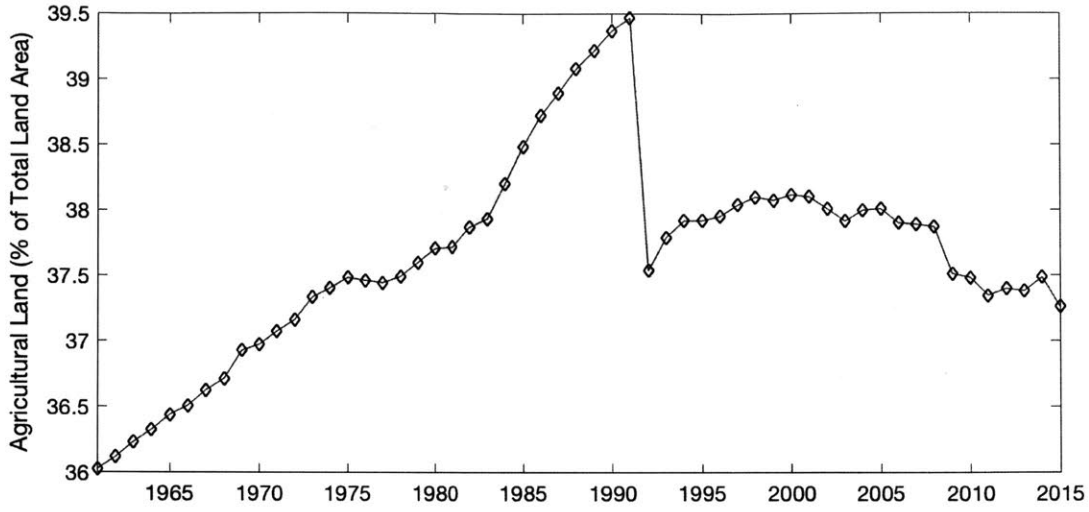


Figure 1-2: Percent of land area occupied by agricultural land shown from 1961-2014 (Figure Source: World Bank, World Development Indicators (World Bank, 2017). Data Source: Food and Agriculture Organization)

more acute issue for short-term food security. The FAO estimates that nearly 800 million people globally were undernourished in 2015 (FAO et al., 2015). While this number has decreased from previous decades, issues of food security are made more difficult to manage with the co-existence of economic and political instability in many of the most vulnerable regions (FAO & WFP, 2017). Although a definite link between climate-change-induced impacts on agriculture, water resources, and vulnerability to conflict has not yet been established, it is an intense area of research with wide reaching consequences regarding the connections between food and water insecurity, economic, social, and political stability, and mass migration (Feitelson & Tubi, 2017; Nordas & Gleditsch, 2007; Reuveny, 2007).

Agriculture productivity is tied to climate and to the hydrologic cycle in a more intimate way than many other biomes. Precipitation amounts and timing, soil moisture levels, and temperature trends that cause shifts of several days in the length of the frost and growing season can have significant effects on the yield of a crop (Sacks & Kucharik, 2011). However, this relationship is not one-way. Alternative vegetative growth can change the evaporative regime in a region, and causes a resultant shift in the surface energy balance as this study will show. Irrigation can be used to overcome natural deficits, but in turn introduces a new soil moisture and transpiration allowance. The extent of this feedback loop is vital for understanding regional climate change, and what conditions may exist for agriculture in the future. Therefore, it is reasonable to not only ask how crops will fair in different climate conditions, but also how crops and agricultural development can impact the climate on land

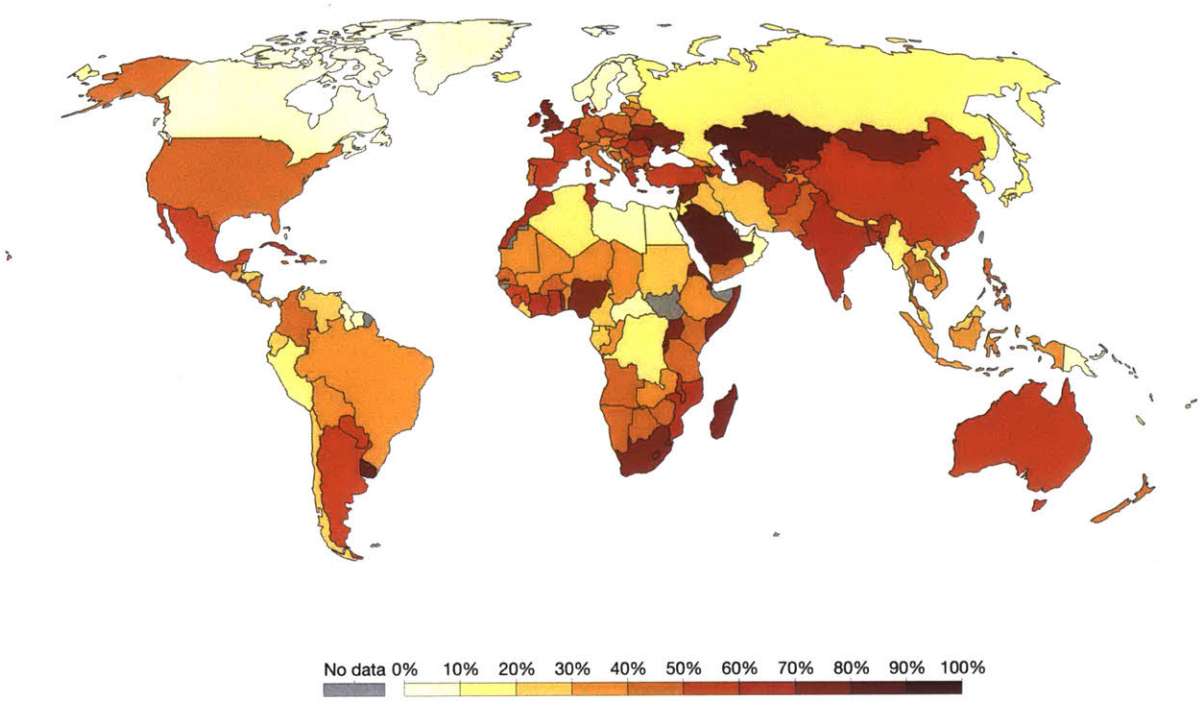


Figure 1-3: Agricultural land as a percent of total land by country (Figure Source: Roser & Ritchie, 2018. Data Source: World Bank, World Development Indicators (World Bank, 2017). Food and Agriculture Organization).

use scales over long periods of time; this is the driving question behind this study.

This study begins with a review of the region of interest — the Midwestern United States — and the proposed connections between land use and climate found in the literature. The MIT Regional Climate Model (MRCM) is then described and calibrated with a focus on matching the patterns of agricultural productivity that were observed in this region over the 20^{th} Century. Using a set of evolving long-term simulations, the analyses performed here will attempt to describe the relative contribution of changing atmospheric composition and land use change (irrigated and non-irrigated cropland development) to the observed regional climate change in the Midwest from 1900 to 2005.

1.1 Background and Study Focus

The Midwestern United States — encompassing Illinois, Indiana, Iowa, Kansas, Michigan, Minnesota, Missouri, Nebraska, North Dakota, Ohio, South Dakota, and Wisconsin in this study — was chosen because it has experienced GHG increases but also dramatic land use change (Figure 1-4). An increase in cropland extent and a dramatic increase in crop yield

over the 20th century was accompanied by a less dramatic but still substantial increase in irrigated areas (Siebert et al., 2015), allowing us to study the relative effects. Additionally, the increases in productivity and yield in this region have been spatially similar, allowing a generalized description of trends in this study (See Chapter 4).

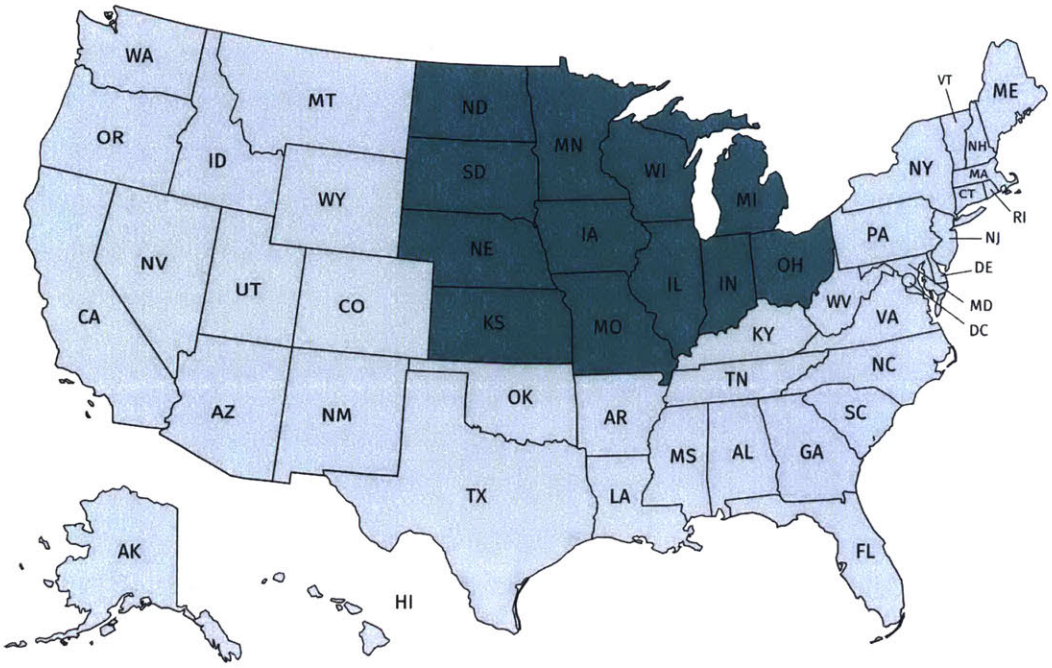


Figure 1-4: Map of the Midwest

The Midwest as a region has already begun to experience the effects of climate change, and further changes are predicted for the future. Crops have seen changes in growth patterns through increased CO_2, and longer growing seasons due to more frost-free days (Melillo et al., 2014). There is a consensus that patterns of precipitation are shifting, with the Midwest experiencing heavier rainfall events, and a higher frequency of the heaviest rainfall events (Melillo et al., 2014). This change in precipitation patterns is expected to continue (Ashfaq et al., 2016). While overall trends throughout the century tend towards warming, studies of observational data in the Midwest have shown a cooling and wetting trend, and a period of significant change in the climate regime from approximately 1950-1970 (Alter et al., 2015a; Douglas, 2016). Douglas (2016) identified a Region of Significant Change (ROSC) (39 – 48° N and 100 – 82° W) and showed that a comparison of 1920-1949 to 1970-1999 climatologies identified a period of significant change. This area and comparison period will be used in later analysis.

An agriculturally predicated cooling in the Midwest during the mid-20th century would in no way contradict the warnings of climate change effects that come from the international scientific community. Anthropogenic climate change builds on a natural system that has caused warming and cooling of the Earth throughout the millennia, and natural and man-made effects combine in ways that are regionally and locally unique. Instead, this study should call even greater attention to the idea that human activities, even on decadal scales, can have significant impacts on regional and local climate. Additionally, as patterns of agricultural expansion and intensification change in the future, the balance between chemically induced atmospheric changes and land use impacts will shift, and the trends of previous decades may not be applicable to future projections. A continuously evolving climate summary that is able to approximate the relative impacts of land use and greenhouse gas changes in the region is vital to understanding what may happen in the future and how relevant past trends may be.

This study is particularly concerned with Midwestern agricultural areas and their relationship to climate because of their importance within the global food production network. This area, known colloquially as the Corn Belt, is an area of high productivity for corn as well as soybean, but is also the area with the world's highest interannual corn yield variability (Ben-Ari & Makowski, 2014). It produces approximately 40% of the world's corn on 20% of the allocated cropland, but extreme years have shown a yield variability of 30% to 50% relative to the mean (Ben-Ari & Makowski, 2014; Hicke & Lobell, 2004; Kucharik & Ramankutty, 2005). As population and food demand increase, agricultural areas must become more productive, and high productivity areas for corn and soybean, such as the Corn Belt, must continue to improve (Ray et al., 2013). Therefore, an understanding of the interaction between agriculture and the climate in this region, now and into the future, is vital for long term planning.

1.1.1 Agricultural Expansion in the Midwest

Despite a decline in the total amount of agricultural land in the United States since the 1950's, it continues to be a substantial portion of land use, with cropland, pastureland, and grazing land accounting for more than 50% of the 2.3 billion total acres in the U.S (Nickerson & Borchers, 2012). In 2007 there were 408 million acres of cropland (17.8% of total land area), and in 2015, U.S. farm output totaled $136.7 billion and accounted for 1% of gross domestic product (ERS, 2016; Nickerson & Borchers, 2012). The Midwest has the highest proportion of land devoted to cultivation of any region in the United States – approaching 50% (Melillo et al., 2014).

This study focuses mainly on the data and characteristics associated with corn and soybean. In most states, harvested corn acreage has been steady or declining through the last century, with the notable exceptions of Illinois, Iowa, and Minnesota which have seen increases in excess of 4 million acres (USDA-NASS, 2016). Harvested soybean acreage, on the other hand, has been increasing consistently in most Midwestern states since the 1920's when survey data is first available from the USDA, counter to the overall cropland trend. Series' of harvested acreage by state are available in Appendix C. A dataset to capture the spatial changes in cropland location throughout the 20th century will be discussed in Chapter 3.

1.1.2 Agricultural Intensification in the Midwest

While the spatial extent of cropland has increased in some areas over the past century, the main change has been the massive increases in yield, spurred by an increase in productivity (Hicke et al., 2004). Corn has seen a weighted average increase among all states of 370% from 1900 to 2016, with almost all the increase occurring after 1935 (Figure 1-5). Soybean has seen a weighted average increase among all states of 400% from 1924 to 2016 (Figure 1-6). USDA-NASS survey data is not available for soybeans prior to the mid 1920's but soybeans have experienced increased yield from this point onward, also approximately along a linear trend. While the larger trends in both of these crops are positive and linear, the time-series are marked by a large interannual yield variability, a defining characteristic of the Midwest (Ben-Ari & Makowski, 2014). In the corn yield, large year to year variations can be seen in 1983, 1988, and 2012 marking drought periods and in 1993 marking a flood year. These events are seen to a lesser extent in the soybean yield record.

Studies have shown that the dramatic yield increases in corn and soybean have been mainly due to improved management (including fertilization, breeding and irrigation), but also to increases in favorable growing conditions (Duvick, 2005; Kucharik & Ramankutty, 2005; Sacks & Kucharik, 2011). The ability to match these advances or take advantage of changing conditions will determine future yield trends. In order to represent the impacts of agriculture on Midwestern regional climate, parameters must be chosen in order to represent the dramatic increases of yield through time. Chapter 4 will include a discussion of the selection of net primary productivity as a modeling stand in for yield, and the validation of its use through observational data.

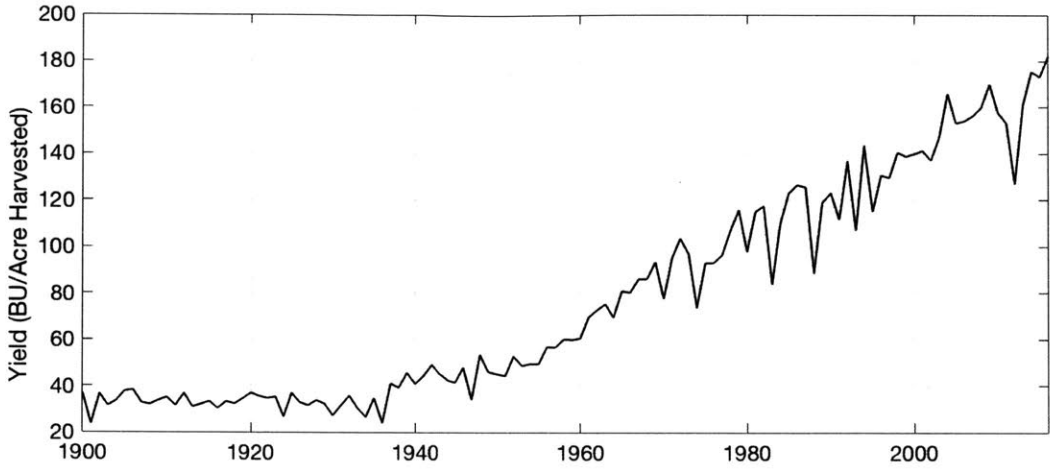


Figure 1-5: Weighted average Midwestern corn yield (Data Source: USDA-NASS, 2016)

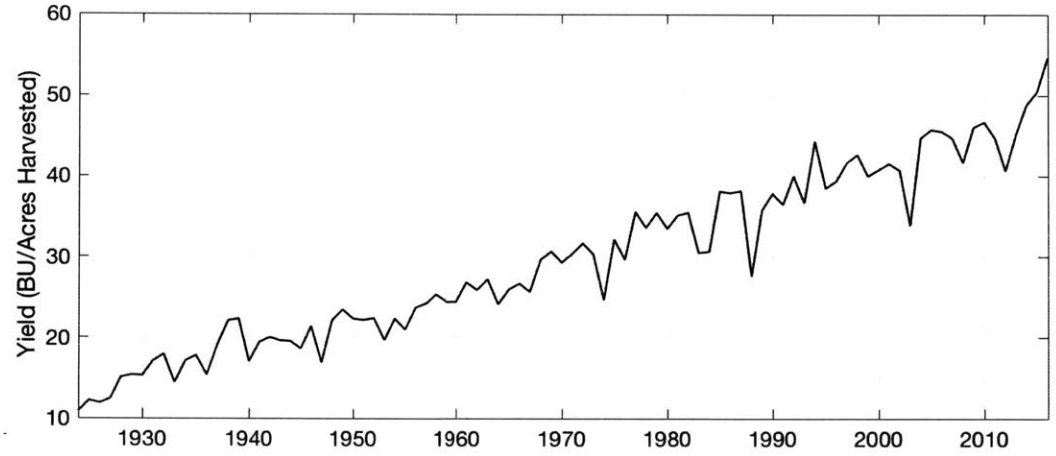


Figure 1-6: Weighted average Midwestern soybean yield (Data Source: USDA-NASS, 2016)

1.1.3 Irrigated Agriculture in the Midwest

Another component of agricultural evolution in the United States is the expansion of irrigation, which started intensively at around the same time that crop yields began to increase (Siebert et al., 2015). In 2013 there were approximately 56,000,000 acres of irrigated farmland in the United States, with 15,760,000 acres in the Midwest (USDA, 2014). Many past irrigation studies of the United States have focused on California, which had early and intense irrigation development throughout the 20^{th} century (Bonfils & Lobell, 2007; Jin & Miller, 2011; Kueppers et al., 2007; Lo & Famiglietti, 2013). The proportion of irrigation growth has been much smaller in the Midwest in the context of its overall cropland area, but is still substantial. While the focus is mostly on agricultural development and climate change

within the Midwestern U.S., there are several other important irrigation areas in the Texas panhandle and the Mississippi River Valley that are included within the model domain.

Advancements in irrigation technology have been aimed at increasing water use efficiency, moving from furrowed irrigation to center pivot systems, and now to low pressure sprayers and drip irrigation systems (Colaizzi et al., 2006). Increased sophistication in irrigation scheduling has also led to improvements in application. While crop area irrigated went up 950 thousand total acres (+1.5%) from 2005 to 2010, water withdrawals for irrigation decreased by 9% to 115,000 million gallons per day; this is significant as irrigation withdrawals in 2010 accounted for 38% of all water withdrawals (Maupin et al., 2014).

Irrigation in many parts of the world allows crops to be grown in arid environments that would not be able to naturally support the crop. In the U.S., irrigation helps to improve yields but can also decrease inter-annual yield variability (Kucharik & Ramankutty, 2005). Irrigation is included in this study as a factor that impacts the water and energy budget in localized areas. A recent dataset for the fractional area of irrigated land will be discussed in Chapter 3.

1.2 Literature Review

As mentioned before, the aim of this study is to characterize the relative effects of a global and a regional forcing – GHG and land use change respectively – in a region that has experienced rapid evolution of both over the 20th century – the Midwest. Several studies are reviewed here that discuss the issues inherent to land use change in climate research, particularly with regards to agriculture (both irrigated and non-irrigated).

1.2.1 Land Use & Land Cover Change

Land use and land cover are distinct ways to characterize environments using both their natural and man-made characteristics. Land cover relates mostly to the physical vegetation characteristics in any given biome, and is liable to change with natural processes such as desertification or burning from wildfires. Land use change refers to the ways that humans change the natural environment and vegetation present in a region in order to make it suitable for their own purposes. Land use change involves conversion into cropland, pasture and grazing land, as well as urbanization (Melillo et al., 2014). As land surface models have become more complex, so has researchers' ability to include detailed representations of land cover and land use change and its evolution, and, therefore, it has begun to be included as an important and highly variable climate forcing within large scale climate reviews (IPCC,

2014; Seneviratne et al., 2010).

The impact of LULC changes on climate is a problem that is distinctly suited for study at a regional scale (Bounoua et al., 2002). The variability in characteristics associated with LULC change happen at sub-grid scales relative to a typical global climate model resolution. The higher resolutions inherent to regional climate models allows the representation of several different vegetation types in a space that would have been represented by one dominant type in a global climate model (Malyshev et al., 2015; Ramankutty & Foley, 1999). It is important to accurately capture the historical evolution of vegetation in an area as well. Studies into the impacts of vegetation have shown that the magnitude and nature of the climate effects depend on both the natural and post-development vegetation in a given area, as well as the location of study (Bounoua et al., 2002; Halder et al., 2016; Twine et al., 2004).

This individuality in the response of various regions highlights the importance of conducting targeted and long-term climate studies tailored to the domain of interest. Additionally, several studies suggest that rapidly changing land use conditions in the 20^{th} century may have partially masked the signal of GHG triggered climate change in the Midwest, (Bonfils & Lobell, 2007; Kueppers et al., 2007) or that the effects have been intimately coupled (Alter et al., 2015a). Therefore, it is important to understand the relationships mentioned above so that LULC can be robustly integrated into the modeling of future climate scenarios, and serve as a concurrent forcing mechanism as patterns of human settlement and land use continue to evolve in the future.

Although there has been past work looking at the climate impact of vegetation changes in the Midwest region, (Alter et al., 2015a; Diffenbaugh, 2009; Mueller et al., 2016; Twine et al., 2004) this study will have several important improvements; namely, the consideration of continuous expansion and intensification of agriculture throughout the 20^{th} century, rather than a simple sensitivity study within a single time period. Additionally, few studies have focused on a combination of expansion, intensification, and irrigation in their evolution of LULC in the Midwest. This study will consider agricultural land as it pertains to cropland, without consideration of land dedicated to pasture and grazing.

1.2.2 Crops and Climate

As noted earlier, agricultural land is roughly a third of the total land area on Earth, and roughly half of land in the United States. In the Midwest, agriculture is important because not only has it been introduced densely as well as broadly, but because agricultural areas have overtaken grasslands as well as forested areas (Ramankutty & Foley, 1999). Cropland conversion can impact the hydrologic cycle in unique ways due to existing climatic conditions,

water availability, and crop phenology variations. For example, Sterling et al. (2012) showed that agricultural areas have a higher evapotranspiration rate than grasslands when irrigated and a lower evapotranspiration rate otherwise. Twine et al. (2004), on the other hand, showed that the longer evapotranspiration period of summer crops actually increased annual evapotranspiration versus natural grasslands. The influence of crop productivity in either study is unclear, and irrigation was not considered in Twine et al. (2004). The impact of conversion to rain-fed cropland on the temperature and evaporative regimes of the Midwest will be one of the focuses of this study.

Agricultural expansion alone has an impact due to the replacement of natural vegetation, but increased productivity, another hallmark of Midwest agricultural development, can also cause major changes. Net Primary Productivity (NPP) is a good proxy for crop yield, as will be demonstrated in a later section, and NPP is directly related to photosynthetic activity as shown below:

$$NPP = GPP - Respiration$$

Gross Primary Production (GPP) is related directly to the photosynthesis rate, which determines the growth of a plant through carbon uptake, but also the corresponding loss of water to transpiration (Roxburgh et al., 2005). Research has shown a linear relationship between NPP and evapotranspiration (Rosenzweig, 1968).

It is important to note that over the past few decades the increases in crop yield have been due to a combination of improvements in management, i.e. breeding, fertilization, improved application of irrigation, and not due solely to direct genetic changes in the plant such as increased Rubisco activity and higher leaf photosynthesis rates (Duvick, 2005). However, photosynthesis may be used as a proxy for increased biomass production and higher crop densities in a given region. The methodology by which this was done will be explained in Chapter 4.

Another confounding effect is the fact that many studies have discussed the effect of a changing climate on yields, both over time and interannually. Lobell and Asner (2003) reported that climate has had a measurable impact on corn and soybean, with 25% of corn yields and 32% of soybean yields being explained by temperature increases from 1982-1998. They did not find a link to precipitation. In their attempt to isolate climate effects, Twine and Kucharik (2009) found that 19% (23%) and 11% (44%) of corn and soybean yield trends respectively could be linked to temperature (precipitation) trends from 1982-2002. Ray et al. (2015) found that 42% of the interannual variability in corn yields in the Midwest could be explained by climate variability, and in some counties this figure is closer to 70%. Because of the large spatial extent of the area, variability is not uniform, and Ray et al. (2015) also

found that precipitation-based variability was dominant in the middle of the region while temperature variations dominated in the north. CO_2 can also have a significant impact on crop growth and photosynthesis rates, depending on the plant (Twine et al., 2013). Finally, Sacks and Kucharik (2011) explored the yield boost that has occured in conjuction with slightly lengthened growing seasons in the Midwest.

It is important to understand the impact of climate on crop yields, as this may become more important with future regional warming. An appreciation of the feedback that exists between crops and climate is crucial to understanding the impacts of one on the other. Additionally, it has been shown that yield increases over the last century may have actually resulted in crop yields that are more sensitive to stresses imposed by droughts – an important characteristic in a world that needs to dramatically increase food production while also being more likely to be plagued by extreme hydrological events (Lobell et al., 2014).

1.2.3 Irrigation Impacts

In many arid and semi-arid areas of the world, natural precipitation amount and timing is not sufficient to maintain large scale agricultural operations. Irrigation introduces anthropogenic control to the temporal and spatial availability of water in a given is therefore an important component of land-surface models.

Irrigation most directly affects soil moisture at the site of application and, through this, the hydrologic cycle. Evapotranspiration and latent heat flux, a measure of the turbulent transport of moisture away from the surface and into the atmosphere, are increased over irrigated areas (Adegoke et al., 2003; Cook et al., 2015; Eltahir, 1998; Harding & Snyder, 2012; Huber et al., 2014; Lo & Famiglietti, 2013; Ozdogan et al., 2010; Qian et al., 2013). Sensible heat flux, the complementary component of turbulent heat transfer away from the surface, is reduced as a result. This increase in latent heat flux is accompanied by surface cooling in irrigated areas (Huber et al., 2014; Jin & Miller, 2011; Kueppers et al., 2007; Lobell et al., 2008; Mueller et al., 2016; Qian et al., 2013). Jin and Miller (2011) found that irrigation promoted cooling in both T_{max} (daytime) and T_{min} (nighttime) temperatures in a study of the California Central Valley, but that nighttime cooling occured only when irrigation was significantly intense enough to overcome warming effects from conversion to cropland. However, the study did not analyze the impacts of greenhouse gas related warming effects that might have influenced observed increases in T_{min}.

The second documented effect of irrigation is an increase in precipitation downwind of the irrigated area, to a greater degree and even in the absence of precipitation over the area itself (Alter et al., 2015a, 2015b; DeAngelis et al., 2010; Huber et al., 2014; Im & Eltahir, 2014; Im

et al., 2014; Lo & Famiglietti, 2013; Lu et al., 2017). These effects are tied to an alteration of the large-scale circulation, moisture transport, and convective environment (Alter et al., 2015a, 2015b; Cook et al., 2015; Halder et al., 2016; Im et al., 2014; Kueppers et al., 2007; Lu et al., 2017; Pal & Eltahir, 2002). Correspondingly, decreases in planetary boundary layer (PBL) height have been found due to modification of energy and moisture fluxes, although studies note that there is uncertainty regarding the ultimate impact on convective potential directly over irrigated areas due to competing effects (Harding & Snyder, 2012; Im & Eltahir, 2014; Im et al., 2014; Jin & Miller, 2011; Lu et al., 2017; Qian et al., 2013). Several studies have noted that circulations induced by cooling over irrigated areas can either augment or counter existing atmospheric patterns such as the Lower Level Jet (LLJ) (Huber et al., 2014) or monsoonal circulation (Im et al., 2014), and influence the overall pattern of moisture advection and precipitation change.

Also of interest in the area of irrigation-induced precipitation is the unique mesoscale circulation pattern that occurs in hot spots of sharp moisture gradients (Frye & Mote, 2010; Lu et al., 2017). Frye and Mote (2010) noted that there is a soil moisture gradient critical value that exists based on background conditions (including the presence of a LLJ) that, once reached, enhances convection along the soil moisture boundary. One of these "hotspots" was identified in the Midwest by Koster et al. (2004). However, it is noted in the paper that only half of the models in the study identify this region as a hot-spot, and Green et al. (2017) questions this designation for this region; they assert that other factors such as large-scale moisture convection may be more strongly linked to precipitation.

Finally, irrigation is a focus not only because of the past work that has been done showing it to be a vital component of land use induced climate changes, but also because it plays a key role in a world that will potentially become drier. Due to overall warming temperatures and an uncertain distribution of precipitation patterns in the mid-latitudes (Melillo et al., 2014), there is considerable debate about the future of soil moisture trends and resulting droughts in areas such as the Midwest (Dai, 2012; Diffenbaugh & Ashfaq, 2010; Mishra et al., 2010; Winter & Eltahir, 2012a). Irrigation is a solution to combat increasing soil moisture deficits, but will require a broader conversation about water withdrawals and consumption in water stressed areas (Melillo et al., 2014).

Chapter 2

Model Structure and Modifications

2.1 MIT Regional Climate Model (MRCM)

Regional Climate Models (RCMs) have the advantage of higher spatial resolutions than Global Climate Models (GCMs), allowing for analysis of climate changes on regional and local scales. While GCMs have grid resolutions that are typically in the hundreds of kilometers, RCMs are able to simulate results at resolutions in the tens of kilometers (Bhuvandas et al., 2014). This higher resolution of RCMs helps to address one of the largest problems with accurate climate modeling — parameterization. Parameterization involves the representation of physical processes that occur at sub-grid scales, such as cloud formation and convection (Peixoto & Oort, 1992). Reduction of grid cell size translates to a smaller area over which conditions are homogenized. However, computation requirements at higher resolutions are often more demanding depending on the size of the domain, and require the appropriate downscaling methods when RCMs are nested in GCMs (Bhuvandas et al., 2014). With current computational capacities, RCMs are valuable tools for climate analysis and are vital for determining the often-unique local impacts of climate change.

The simulations for this study were performed using the MIT Regional Climate Model (MRCM). MRCM is an updated version of the Regional Climate Model Version 3 (RegCM3) climate model, originally developed at the National Center for Atmospheric Research (NCAR) and maintained by the International Center for Theoretical Physics (ICTP). The model has been further modified by the Eltahir research group to improve representations of albedo, dust emissions, cloud and convection schemes, and boundary layer dynamics (Gianotti & Eltahir, 2014a, 2014b; Gianotti et al., 2012; Marcella & Eltahir, 2010, 2012). MRCM is composed of an atmospheric circulation model coupled with a land surface model. The present setup uses the Grell Cumulus Scheme with the Arakawa and Schubert convective closure assumption (Arakawa & Schubert, 1974; Grell, 1993). This combination has been

shown as the most appropriate in previous studies of the Midwestern United States (Winter, 2006, 2010; Winter & Eltahir, 2012a). Further description of MRCM components can be found in Marcella (2013) and Winter (2006).

MRCM has a standard resolution of 30 km horizontal grid increments and 18 vertical sigma levels from the surface to a top pressure of 50 hPa. The domain is centered on 40.5° N, 91.5° W for this study, with 122 zonal elements and 80 meridional elements in the grid. The selection of this domain allows for analysis of the Midwest and surrounding regions, without incorporating much of the mountainous regions in the West. The simulations will be run from 1900-2005 or 1901-2005 depending on the reanalysis data used, with a January 1 start date for all simulations. This time period is useful because of the large increase in agricultural productivity that occurred in the Midwest during the later portion of the 20th century. This time period also captures the spatial expansion of cropland that occurred before it reached relative stability in mid-century (Ramankutty & Foley, 1999). A full description of the experimental design is available in Chapter 5.

2.2 Integrated BIosphere Simulator (IBIS)

2.2.1 Structure

A vital development for MRCM was the replacement of the original land surface model, the Biosphere-Atmosphere Transfer Scheme (BATS), with the Integrated BIosphere Simulator (IBIS) (Winter, 2006). Several studies have noted the suitability of RegCM3 (Diffenbaugh, 2009; Diffenbaugh et al., 2005; Kueppers et al., 2007), and of a coupled RegCM3-IBIS framework, for regional climate modeling in the Midwestern United States (Winter & Eltahir, 2012a, 2012b).

IBIS calculates the full energy, water, and carbon budgets and the feedbacks that they produce (Foley et al., 1996; Kucharik et al., 2000). The momentum, thermal energy, and water budget equations included in IBIS are largely based on the land-surface-transfer scheme (LSX) which then works in tandem with the other IBIS components (Pollard & Thompson, 1995). In the documentation of the coupling of IBIS to RegCM3 as a replacement to BATS, Winter (2006) notes that the addition of dynamic vegetation capabilities, increased soil layers, more complexity in PFT combinations within single grid cells, and generally improved parametrizations of plant physiology among other things led to the decision to update the model. This model, part of the "third generation" of land surface models that include both a consideration of stomatal resistance and more complex photosynthesis modeling, have greatly improved researchers' ability to accurately model land-atmosphere interactions (Seneviratne

et al., 2010). Despite this increased complexity, studies have still called into question the ability of lumped models to accurately represent these interactions in the face of fine-scale variations in parameters that cannot be fully captured with average attributes (Wood, 1997). This parameterization issue is similar to that dealt with in the representation of atmospheric processes.

When IBIS and RegCM3 were first coupled, the model underestimated latent heat flux and consequently evapotranspiration due to an excess absorption of solar radiation, especially during the summer (Winter, 2006). This was later fixed through the adoption of a new convection scheme closure and an update to water vapor properties (Winter & Eltahir, 2010), although extra care should be taken to ensure that these measures are accurately represented in the model in all scenarios. A study by Cunha et al. (2013) of caatinga vegetation in Brazil with an offline version of IBIS also shows difficulty in measuring the annual variability of latent heat flux in response to moisture and vegetation changes, although the setup of the study makes the comparison to the fully coupled model less ideal (Cunha et al., 2013). Study of LSX alone, the base land-surface scheme for IBIS, has shown that it can appropriately represent the hydrologic cycle, of which evapotranspiration is a vital component; however, the study was similarly performed with an offline model and therefore could not capture the effects of being coupled to an atmospheric model and the feedbacks that entails (Levis et al., 1996).

IBIS structure accounts for calculations within two vegetative layers, an upper and a lower canopy (Foley et al., 1996). Within the modified version of IBIS there are seventeen biome types that are designated through a user generated land surface file, sixteen of which are found within the current study domain. There are also thirteen Plant Functional Types (PFTs) that can exist to varying degrees and in varying combinations in a biome, based on both climatic variables and user designated fractions of leaf area index (Foley et al., 1996) (Table 2.1). The PFTs are distributed between the upper and lower canopy, with trees being assigned to the former while grass and shrub type vegetation are assigned to the latter. While IBIS can be set to dynamically update vegetative growth and classification, the static option with an initial land surface boundary condition was used in this study to maintain consistency between simulations.

The main update from the original IBIS biome and PFT catalogue was the inclusion of a crop plant functional type, and both non-irrigated and irrigated cropland biome types (Kucharik, 2003; Kucharik & Brye, 2003; Marcella, 2013). The original tundra biome (13) was adjusted to reflect the characteristics of irrigated cropland areas, due to there being no tundra located within the domain. This setup precluded the addition of an entire biome scheme into the model while still allowing for the division between rain fed and irrigated

Table 2.1: Plant Functional Types (PFT) in current setup of MRCM-IBIS. Index 1-12 show original IBIS PFTs (Foley et al., 1996) with the addition of cropland (13)

Plant Functional Type	Index
Tropical Broadleaf Evergreen Trees	1
Tropical Broadleaf Drought-Deciduous Trees	2
Warm-Temperate Broadleaf Evergreen Trees	3
Temperate Conifer Evergreen Trees	4
Temperate Broadleaf Cold-Deciduous Trees	5
Boreal Conifer Evergreen Trees	6
Boreal Broadleaf Cold-Deciduous Trees	7
Boreal Conifer Cold-Deciduous Trees	8
Evergreen Shrubs	9
Cold-Deciduous Shrubs	10
Warm (C4) Grasses	11
Cold (C3) Grasses	12
Crops (C4)	13

cropland that exists in the Midwestern United States. The crop PFT is only found in either non-irrigated or irrigated biomes, and these biomes contain no other vegetation type besides cropland. Additionally, the physiology of the crop plant functional type is based on corn, a C_4 photosynthesis pathway crop. This leads to a set of necessary assumptions in representing Midwestern agricultural development that are discussed in Chapter 4. The transition from the original IBIS biome setup to the current setup can be seen in Table 2.2.

IBIS is structured with six soil layers that increase in thickness with depth (10, 15, 25, 50, 100 and 200 cm respectively) for a total soil depth of up to 4 m, and that allow for independent calculations of soil temperature, moisture, and ice fraction (Foley et al., 1996). However, in Winter (2010) it was determined that the unsaturated depth in the Midwest, and particularly for a test area in Illinois, rarely reached a depth of four meters and the sixth layer was disengaged, leaving a composite soil depth of 2m across five layers. Additionally, the model currently uses a free drainage assumption, where a user defined coefficient can allow either an impermeable bottom layer (bperm = 0) or a fully permeable bottom layer (bperm = 1). The default value in the model is set to 0.1. Yeh and Eltahir (2005) noted the problems with the free drainage model, especially in shallow water table areas such as Illinois, where feedback from the aquifer should be modeled in order to accurately represent subsurface storage.

Irrigation was initially introduced into MRCM for testing land-atmosphere processes in semi-arid regions, and the original irrigation scheme was set to return the root zone soil

Table 2.2: Comparison of the original IBIS biome configuration (Foley et al., 1996) and the current setup used in MRCM

Original IBIS Biomes	Modified IBIS Biomes	Index
Tropical Evergreen Forest	Tropical Evergreen Forest	1
Tropical Deciduous Forest	Tropical Deciduous Forest	2
Temperate Evergreen Broadleaf Forest	Temperate Evergreen Broadleaf Forest	3
Temperate Evergreen Conifer Forest	Temperate Evergreen Conifer Forest	4
Temperate Deciduous Forest	Temperate Deciduous Forest	5
Boreal Evergreen Forest	Boreal Evergreen Forest	6
Boreal Deciduous Forest	Boreal Deciduous Forest	7
Mixed Forest / Woodland	Mixed Forest / Woodland	8
Savanna	Savanna	9
Grassland / Steppe	Grassland / Steppe	10
Dense Shrubland	Dense Shrubland	11
Open Shrubland	Open Shrubland	12
Tundra	Irrigated Cropland	13
Desert	Desert	14
Polar Desert / Rock / Ice	Polar Desert / Rock / Ice / Tundra	15
Ocean	Non-Irrigated Cropland	16
Inland Water	Ocean and Inland Water	17

moisture at every time step to the weighted average root zone field capacity in each of the four top layers that make up the root zone (0-100 cm) (Marcella, 2013). The same irrigation setup was later used to investigate the effects on precipitation in the Gezira region in East Africa (Alter et al., 2015b) and in West Africa in conjunction with the West African monsoon (Im & Eltahir, 2014; Im et al., 2014). In the current setup, the model is slightly modified in order to more realistically represent irrigation in the region. Rather than replenishing the root zone constantly, irrigation is applied when a threshold of 75% of average root zone relative field capacity is reached. Additionally, irrigation is restricted to July – September rather than May – September, which aligns with a restriction on crop growth before July.

Chapter 3

Initial and Boundary Condition Data

The scope of the simulations in this thesis was made possible through the availability of new historical data sets spanning the 20th century. The data are used as initial and boundary conditions for the domain, and can express long term trends as they evolve throughout the decades. Many of the boundary conditions use either recently released data or combine multiple data sources in a way that is most representative of certain land use conditions in the Midwestern United States throughout the 20th century. This chapter describes all data sets used, as well as greenhouse gas time series. Any observational data sets that were used for selection and validation of internal model parameters to represent productivity (yield and NPP) are described in Chapter 4.

3.1 Reanalysis Products

For historic simulations, MRCM requires the use of boundary conditions on the lateral and surface boundaries of the domain. The recent development of reanalysis data spanning the 20th century provides lateral boundary conditions, and allows forcings to be paired with contemporary climate conditions. Reanalysis, or "retrospective analysis", datasets are made by assimilating observational data into a numerical model that provides short-term first guess forecasts. The benefits of reanalysis data include a representation of climate conditions that are spatially uniform and physically consistent, and that provide more descriptors of the climate system than are available from observations alone (Buizza et al., 2017). For all of the reanalysis data used, the model requires information for air temperature, relative humidity, geopotential height, and zonal and meridional winds obtained at the synoptic times 00:00, 06:00, 12:00, and 18:00 GMT.

3.1.1 ERA20C

ERA20C was the first 20th century atmospheric reanalysis product produced by the European Centre for Medium-Range Weather Forecasting (ECMWF) as a part of the ERA-CLIM project. It spans 1900-2010, and was produced using a 24-hr 4D-Var analysis method on 91 model levels with a horizontal resolution of 125 km, and is composed of a single ensemble member (Poli et al., 2016). The reanalysis is created using an atmospheric general circulation model that was validated with 10 ensemble members and is forced with observations of sea surface temperature and sea ice cover (HADISST2.1.0.0), atmospheric composition changes (ozone, aerosols, and greenhouse gases), and solar forcings (Hersbach et al., 2015). This model-only product is available from ECMWF as ERA-20CM. All ECMWF datasets described here can be downloaded from their public interface at https://www.ecmwf.int/en/forecasts/ datasets/browse-reanalysis-datasets.

The reanalysis dataset is built from the assimilation of atmospheric surface pressure data from the International Surface Pressure Databank (ISPD 3.2.6) and the International Comprehensive Ocean-Atmosphere Data Set (ICOADS 2.5.1). Marine winds are also used from ICOADS 2.5.1. The number of observations rises dramatically over the period of the reanalysis, from 30,000 surface pressure observations in 1900 to 3.6 million in 2010 (Poli et al., 2016). ERA20C was a first attempt by ECMWF to assimilate surface observations into a long-term analysis with a new assimilation method (Poli et al., 2013). Poli et al. (2016) notes that there are some deficiencies in the product, such as difficulty capturing precipitation anomalies in the first quarter of the century and a slight negative impact on low-frequency variability trends relative to the model-only product. Stickler et al. (2015), notes that the omission of early century pressure readings associated with hurricanes may have impacted accuracy.

3.1.2 CERA20C

CERA20C is the follow-up dataset to ERA20C and ERA-20CM and was produced by ECMWF as a part of the ERA-CLIM2 project (Laloyaux et al., 2016). It currently represents the most recent available reanalysis dataset covering the full 20th century and is the first coupled atmosphere-ocean 20th century reanalysis produced by ECMWF. It spans 1901-2010 and is produced using the Coupled European Reanalysis (CERA) Assimilation System with an atmospheric resolution of 125 km and 91 vertical levels, a land resolution of 125 km with 4 soil layers, and an ocean grid resolution of 110 km with 42 layers (ECMWF, 2017). CERA20C is produced across 10 model ensembles, and because an ensemble mean was not available at the time of acquisition, member 5 was chosen as a representative member based

on July-August average temperatures over a subset of the Midwest.

Like ERA20C, CERA20C assimilates surface pressures and marine winds, but also ocean temperature and salinity profiles. HADISST2 is used as the SST dataset, providing temperatures at the air-sea interface. Surface pressure data is from ISPD 3.2.6 and ICOADS 2.5.1, and marine winds are also obtained from ICOADS 2.5.1 (Laloyaux, 2017).

3.1.3 NOAA 20CR V2c

The first reanalysis dataset for the 20th century, NOAA 20CR, was developed by the National Oceanic and Atmospheric Administration (NOAA) in collaboration with the Cooperative Institute for Research in Environmental Sciences (CIRES). The dataset involved the use of an Ensemble Kalman Filter method and the assimilation of surface pressure observations only into an atmospheric model forced using SST and sea ices concentrations (HASISST1.1) (Compo et al., 2011). The current version, V2c, spans 1851-2014, includes observations from ISPD 3.2.9, and was generated from 56 model ensembles on a horizontal resolution of 210 km and on 28 vertical model levels (Compo et al., 2015). Additionally, V2c includes sea ice data from COBE-SST2 and sea surface temperature data from the Simple Ocean Data Assimilation with sparse input (SODAsi.2) (NCAR, 2017).

3.1.4 Reanalysis Comparison

The biggest difficulty in generating long term reanalysis data is the lack of available and reliable observations in the early 20th century before the advent of radiosonde and later satellite derived measurements. Additionally, areas such as the Southern Hemisphere that have lower record density even in the latter portion of the century may face accuracy issues (Laloyaux, 2017; Poli et al., 2013). The reanalysis datasets discussed here attempt to maintain consistency in data sources, and therefore utilize only select surface observations. For example, the lack of early data has been shown to be a potential culprit behind the poor representation of winter high- and low-frequency atmospheric variability in the Northern mid-latitudes in ERA20C and NOAA 20CR V2c (Dell'Aquila et al., 2016). A recent study from Stryhal and Huth (2017) looked at the classification of atmospheric circulation patterns and found significant differences in classifications between reanalysis data that led them to suggest that the early portions of these datasets (ERA20C and NOAA 20CR V2c) should be used with caution.

Throughout the course of the study, ERA20C was chosen as the initial reanalysis dataset for conducting the simulations. This was partially due to the extensive use of a previous ECMWF product with MRCM (ERA Interim) but also because of its higher resolution

compared to NOAA 20CR V2c. Ultimately, a second set of simulations was run with the newer CERA20C due to a troubling feature in the surface temperature trend within the Region of Significant Change (ROSC) which is defined in Chapter 5. The results from both of these sets of simulations are analyzed and compared in Chapter 5.

As can be seen in Figure 3-1, ERA20C exhibits a consistent cold bias and an exaggerated cooling trend in the mid-century that begins in 1930. In comparison, the chosen ensemble member of CERA20C has a warm bias that remains relatively consistent across the entire simulation period. While NOAA CR20 V2c most closely matched the observations in most years, it exhibits a cooling trend in the ROSC from 1950 to 1970 that is roughly twice that of the observations for this area.

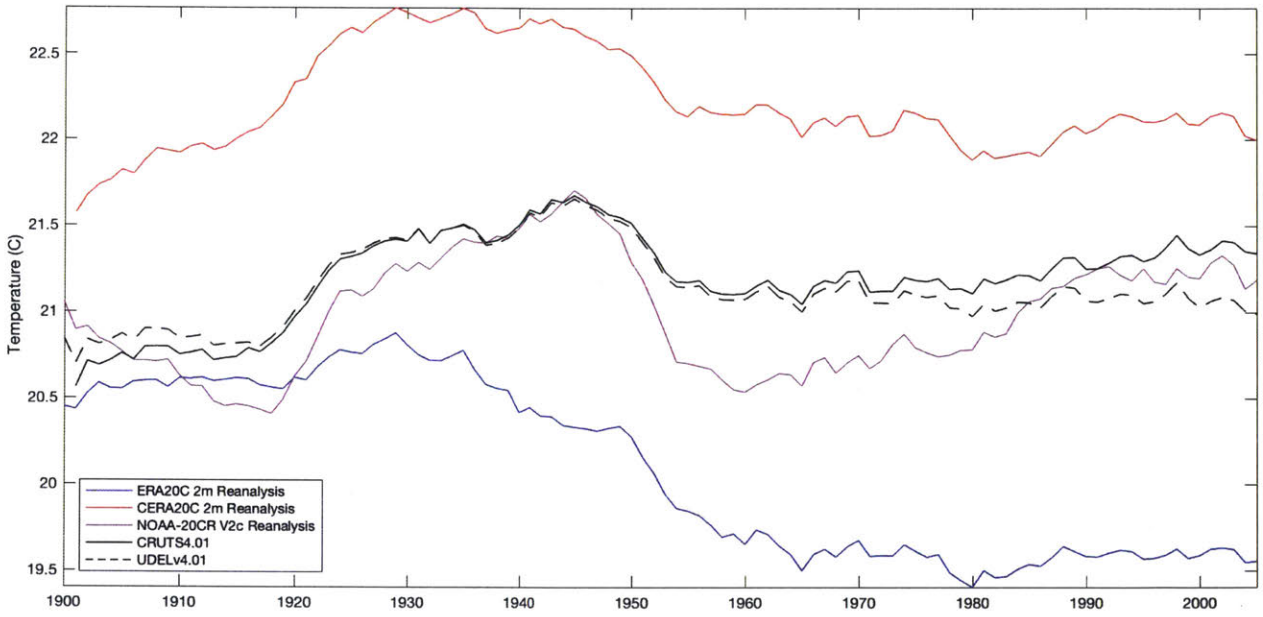


Figure 3-1: Time series of July-August average surface (2m) temperatures plotted in the Region of Significant Change (ROSC) (39-48° North, 100-82° West) from 1900 to 2005 for ERA20C, CERA20C, and NOAA 20CR V2c and compared to CRUTS4.01 and UDELv4.01

3.2 Sea Surface Temperature (SST)

Monthly average sea surface temperature data was used from the Hadley Centre Sea Ice and Sea Surface Temperature Data Set (HADISST v1.1) provided by the Met Office Hadley Centre in collaboration with the National Oceanography Centre, Southampton and the Lamont-Doherty Earth Observatory of Columbia University (Rayner et al., 2003). The dataset provides monthly average sea surface temperature at 1° by 1° resolution from January 1870

to November 2016 at the time of download. The dataset replaces GISST, which has been previously used in conjunction with RegCM3.

Data acquisition and assimilation is described in Rayner et al. (2003). The main data source was the Met Office Marine Data Bank (MDB), which was augmented after 1982 with satellite products from the Global Telecommunications System (GTS). Data from the Comprehensive Ocean-Atmosphere Data Set (COADS) was also used from 1871-1995 to supplement coverage. Additionally, SSTs from the U.S. National Oceanographic and Atmospheric Administration (NOAA) Advanced Very High-Resolution Radiometer (AVHRR) were used from January 1982 onward to supplement deficiencies in the spatial coverage of data. Sea ice data was obtained through digitized sea ice charts and passive microwave retrievals where available.

3.3 Land Use

In order to consider land use and land cover (LULC) change as a forcing within the model, it was necessary to construct a dataset that accounted for cropland expansion and the development of irrigation. Cropland is the only land use change considered — no pasture land, grazing land, or urban land — in order to target the effects of crops specifically. A historical dataset with all necessary components for a progression of years in the 20^{th} century is not available widely, so one was made, while trying to make sure that the combination would be as physically consistent as possible. This required the use of three separate datasets and processing as described below.

3.3.1 Irrigation Data

Irrigation data is obtained from the Historic Irrigation Dataset (HID) which contains global irrigation data in decadal increments from 1900 to 1980 and five years increments from 1980 to 2005 (Siebert et al., 2015). While there are 8 products of the dataset available for download, the set used was HYDE-FINAL-IR, which is the best guess of the distribution of irrigation with a close match to irrigation records while using cropland and pasture extent from the History Database of the Global Environment (HYDE) Version 3.1. The data resolution is 5 arcmin by 5 arcmin (1/12°), and is given in hectares of irrigation per grid cell. The area provided in Siebert et al. (2015) is Area Equipped for Irrigation (AEI), meaning that in any given year the amount of area irrigated may be less than the raw area provided. The area of irrigated land within the full domain increased in the 1950s with the introduction of irrigation in the Central Plains (TX, KS, NE) and again with expansion

along the Mississippi River (Figure 3-2).

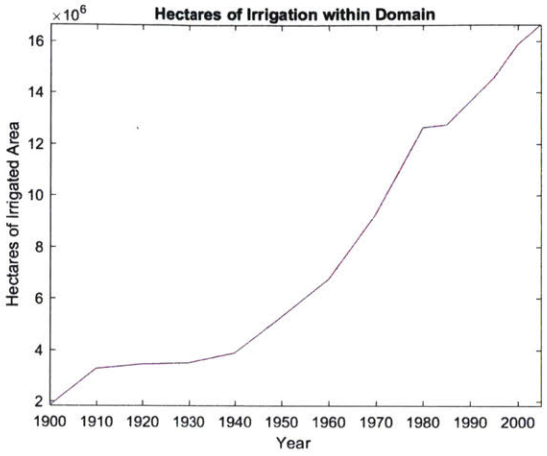


Figure 3-2: Hectares of irrigated area within the domain from decade to decade according to data from the Historic Irrigation Dataset (HID) (Data Source: Siebert et al., 2015)

Siebert et al. (2015) notes that because less data is available, estimations prior to 1960 may be less accurate but that overall, irrigation extent in the United States is more accurate than with previous reconstruction methods, particularly in the central U.S..

3.3.2 Cropland Data

It was also necessary to choose a historical croplands dataset. There are several datasets with long term crop data available but multiple criteria made the Harmonized Global Land Use Dataset (1500 – 2100) (LUH1) the most appropriate for this study (Chini et al., 2014; Hurtt et al., 2011). The dataset is available annually at 0.5° by 0.5° resolution for the full simulation period, and provides the fractional area of cropland coverage for each grid cell. Like the Historical Irrigation Dataset (HID), cropland estimates were developed from a subset of HYDE3.1. This first version was developed as part of the Coupled Model Intercomparison Project Phase 5 (CMIP5) in preparation for the Intergovernmental Panel on Climate Change Annual Report 5 (IPCC AR5). Although this study does not look at future projections, further work will also benefit from the use of this dataset which has data available until the end of the 21st century and is comparable with CMIP5 GCM output that may be used. The amount of cropland within the domain can be seen in Figure 3-3.

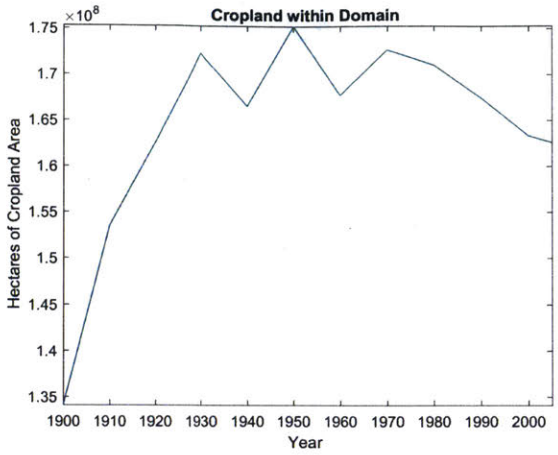


Figure 3-3: Hectares of cropland within the simulation domain according to data from the Harmonized Global Land Use Dataset (LUH) (1500-2100) (Data Source: Chini et al., 2014 and Hurtt et al., 2011)

3.3.3 Potential Vegetation

The irrigation and cropland datasets discussed earlier only denote the areas of irrigated and non-irrigated cropland with no information about other vegetation. Therefore, a potential vegetation dataset developed by Ramankutty and Foley (1999) is used. Potential vegetation is slightly different from natural pre-agricultural vegetation, but describes the vegetation that would likely exist without human influence. The dataset is available at 5 arcmin resolution. Importantly, the land use types used in the dataset are consistent with those defined in IBIS and therefore do not require reclassification or combination. Because this dataset forms the base of the composite land surface dataset, if an area is converted to cropland and then later reverts, the underlying before-and-after potential vegetation is identical.

3.3.4 Composite

The full version is made by layering these three separate datasets according to certain thresholds, and downscaling the cropland dataset to 5 arcmin resolution. If a grid cell is occupied by at least 50% cropland, then the entire grid cell is designated as the non-irrigated cropland biome. If a grid cell contains at least 25% irrigation, the grid cell is designated as the irrigated cropland biome. The final land surface dataset is interpolated by the model in the pre-processing phase in order to match the 30-km grid cell resolution. An example of the final biome map for 1990 can be seen in Figure 3-4. Maps of the biome distribution for all the years used (1900-2000 in ten year increments) can be found in the Appendix (B-6 - B-16)

The static nature of the potential vegetation dataset minimizes the changes that occur

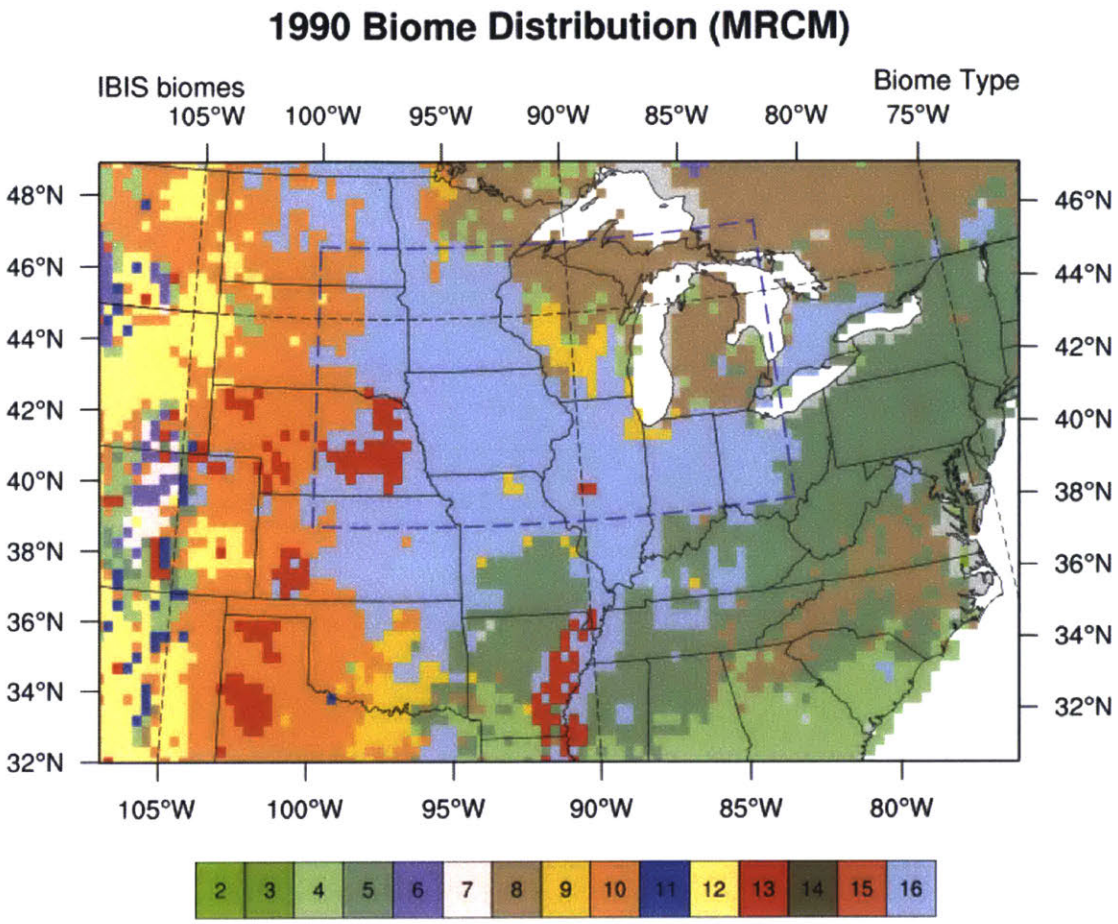


Figure 3-4: Biome distribution after model pre-processing and interpolation to MRCM grid. Full biome indexes can be found in Chapter 2. For reference, non-irrigated cropland (16) and irrigated cropland (13)

throughout the domain from decade to decade and allows the results to be focused on the impact of agricultural development in both irrigated and non-irrigated areas. There are several main areas of change that occur throughout the 20th century between the 11 sequential biome maps that were generated. First is the expansion of non-irrigated cropland north into the upper Midwest and Canada as well as west into the central and lower Great Plains. Cropland also replaces savanna in Missouri and Southern Illinois in the mid-century. However, the majority of cropland area is established by the 1920s and 1930s, as evidenced in the earlier LUH time-series. There are two main areas of irrigation development. A line of irrigation stretching from the Texas Panhandle to Nebraska begins to develop robustly in the 1950s. The second area of irrigation along the Mississippi River is not established until the 1990s and 2000s. In the east, the opposite cropland trend occurs, with small patches of cropland disappearing and transitioning within the model to forested areas. This same

reversal happens in Texas and Oklahoma with the reemergence of savanna, and a reemergence of mixed grassland in Kansas at the end of the 20^{th} century.

3.4 Greenhouse Gases (GHG)

Two of the full-century runs (VG and NVG) are made using real evolution of greenhouse gas concentrations. MRCM utilizes a data-file with a 1765-2005 historical time series detailing the yearly concentrations of 5 greenhouse gases – CO_2 (ppm), CH_4 (ppb), N_2O (ppb), CFC-11 (ppt), and CFC-12 (ppt). The time-series of concentrations during the simulation period can be seen in Figure 3-5. Greenhouse gas concentrations were updated in order to reflect the CMIP5 recommendations made in part in preparation for the Intergovernmental Panel on Climate Change (IPCC) Annual Report 5 (2013) (Meinshausen et al., 2011). This ensures consistency with both intended use of the land use (cropland) data set, and with any future use of CMIP5 component modeling results as comparison or forcing. For the third simulation (NVNG), the concentrations of all 5 greenhouse gases listed above were set to the 1900 values.

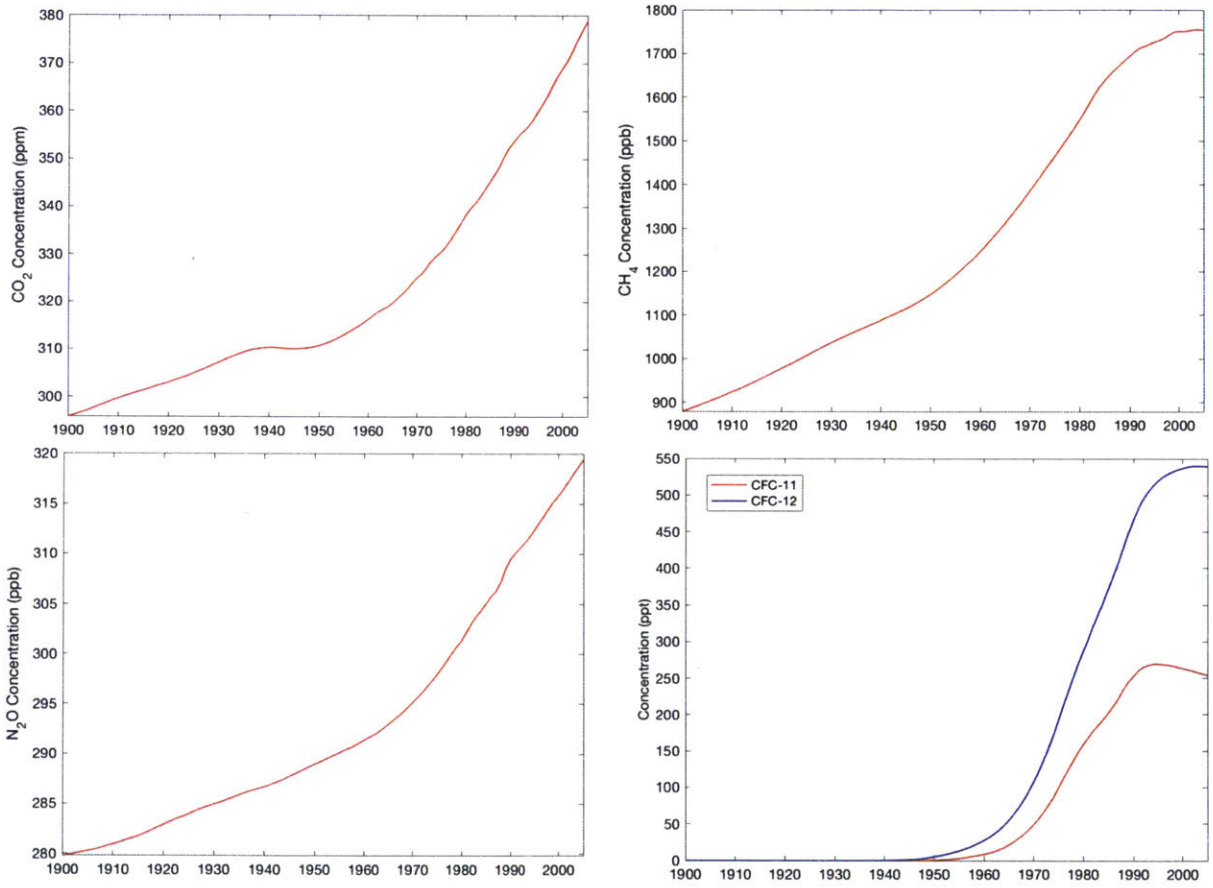


Figure 3-5: Greenhouse gas concentrations from CMIP5 recommendations. CO_2 (top left), CH_4 (top right), N_2O (bottom left), CFC-11 and CFC-12 (bottom right) (Data Source: Meinshausen et al., 2011)

Chapter 4

Representation of Improvements in Agricultural Productivity

Along with expansion of cropland area and introduction of irrigation, an increase in crop productivity has been an important feature in Midwestern agricultural development. Yield increases on the scale that have been observed represent a dramatic increase in biomass and plant activity during the summer months, and this must be taken into account in land use representation. In order to accurately represent yield increases, crop physiology in the model has been calibrated based on observed values of NPP — a value linked to both the metric this aims to represent, yield, and ,through photosynthesis rate and evaporation, to the surface carbon and energy balance.

4.1 Yield Data

Agricultural yield data for corn and soybean were obtained from the U.S. Department of Agriculture's (USDA) National Agricultural Statistics Service (NASS) (USDA-NASS, 2016). Survey, rather than census data, was used in order to get full coverage of the period and data on acres harvested, acres planted, production and yield was collected for each state annually. Production is provided in bushels (BU) and yield in bushels per acre. Data is available since 1900 for corn, and begins in the mid 1920s for soybean. Time series plots of production, acres harvested, and yield for each Midwestern state for corn and soybean can be seen in Appendix C: Corn and Soybean Yield by State (USDA-NASS, 2016). For use in validation, a time series of average Midwestern yield for corn and soybean was created by averaging individual state yield by proportion of production. These figures can be seen in Chapter 1 (Fig. 1-5 & Fig. 1-6).

4.2 Choice of Parameters & Validation Method

4.2.1 Photosynthesis Rate

There are several considerations in order to ensure that the chosen parameter causes the intended vegetation response, while at the same time avoiding physical side effects that are not representative of historical change. A series of assumptions will need to be made in order to match current operational capabilities of the model.

First of all, it must be noted that the yield increases over the past century have been due to a wide range of factors. Much of the improvements in yield over the past century have been due to changes in management practices — pest control, fertilization, and planting efficiency, with planting densities for maize increasing by 1000 plants/hectare/year since the 1950's (Duvick, 2005). Irrigation, as mentioned in Chapter 1, has also contributed to yield increases (Kucharik & Ramankutty, 2005). Duvick (2005) notes that while 50% of the improvement in yield may be due to improved management, another 50% is due to breeding and physiological changes in the plant. IBIS in its current coupled form as a part of the MIT Regional Climate Model does not have a robust mechanism to represent external changes related to management and therefore all of the observed increases in crop productivity must be represented through physiological adjustments to the crop plant functional type within the model.

With this in mind, a parameter is chosen that can be individually modified for crops alone, and that can manifest the changes associated with yield improvements — namely an increase in biomass per area. Net photosynthesis rate is tied in the model to net primary productivity, and will serve as a proxy for increasing vegetative production in an area. While the increase in productivity and yield cannot be ascribed solely to an increase in crop photosynthesis, there is evidence that photosynthesis rate has increased in corn through breeding over the past century, especially in times of plant stress and recovery (Duvick, 2005). Long et al. (2006) notes that keeping everything else constant, increasing leaf photosynthesis would cause an increase in yield. Increasing photosynthesis rate at the leaf level can also serve as a proxy for increased planting densities that did occur (Duvick, 2005).

In MRCM, photosynthesis rates are calculated for each plant functional type, because biomes can have differing PFT compositions even within the same designation. Importantly, IBIS adopts a method of scaling photosynthesis whereby the rates are calculated at the leaf level and then converted to the canopy level using a scale factor based on canopy characteristics such as leaf area index (Foley et al., 1996).

For crops, photosynthetic rate at the leaf level is a function of leaf temperature, soil moisture stress, a leaf respiration coefficient (defined as 0.100 for crops in MRCM), and

maximum rubisco activity. The photosynthesis can be light limited, rubisco limited, and carbon dioxide limited (which is governed by the CO_2 efficiency and the intercellular CO_2 concentration) depending on the relative value of the three. This photosynthesis scheme is representative of a C_4 crop such as corn, and does not model the physiology of C_3 crops such as soybean. Using a comparison of these photosynthetic rates, the model calculates a gross photosynthesis rate which is modified by the dark respiration rate (a function of max rubisco, leaf temperature, and the leaf respiration rate) to obtain the net photosynthesis rate (ancr), where

$$ancr = factor \times (agcr - rdarkcr)$$

This rate can easily be modified by a factor to adjust the net photosynthesis rate for all crops in the domain. The default factor in the model is 1.0. All of these relationships can be seen in the code for IBIS as developed by Foley et al. (1996).

This photosynthesis rate is used, among other things, to determine the crop stomatal conductance and primary productivity, both gross and net. In order to use this parameter to produce realistic representations of net primary productivity, the nature of the relationship between model output of NPP and any multiplication factor that is applied must be understood.

In order to do this, the model was run for two-year simulations from 1991-1992 using the representative land use map for 1992, with the only variation between the models being the photosynthesis multiplication factor. Simulations were run for a multiplication factor of 1, 2, 2.5, 5 and 6 to get a wide distribution of net primary productivity levels. Annual net primary productivity was calculated for 1992 for each of the multiplication factors in each grid cell designated as either non-irrigated cropland or irrigated cropland, and then normalized by the values in the simulation where the factor was equal to six. An average was then taken for all of the grid cells in each cropland type over the domain less the outer nine grid cells to eliminate edge effects. The results show that the model output NPP has a linear relationship with the factor, with R^2 values of 0.99 for both cropland biomes (Figure 4-1).

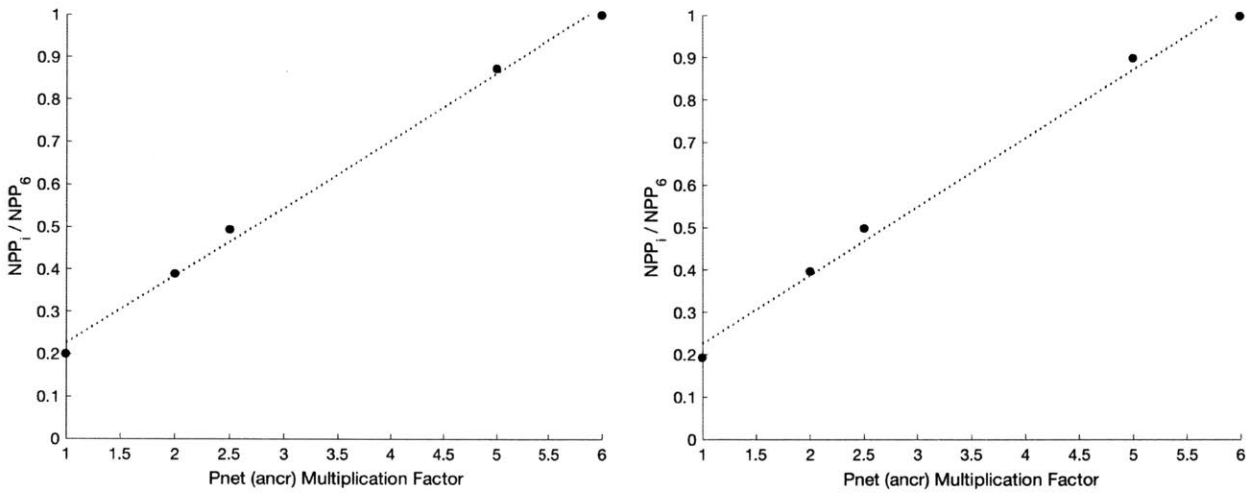


Figure 4-1: Average annual NPP output generated from Pnet multiplication factors. Relationship is plotted for non-irrigated cropland (left) and irrigated cropland (right)

4.2.2 Net Primary Productivity

Net primary productivity (NPP) is a measure of the amount of carbon fixed into new biomass in vegetation, and can broadly be represented as the difference between photosynthetic production and respiration (Roxburgh et al., 2005). However even this definition causes inconsistencies in the ways that NPP is calculated and reported. Roxburgh et al. (2005) notes that there are two common ways that NPP can be presented. In the first NPP is always positive, as a measure of the new material being added to the system. In the second, NPP can be negative in some periods where respiration exceeds photosynthesis. In MRCM, NPP is presented as an instantaneous flux of mol $CO_2/m^2/s$ and therefore fluctuates from negative to positive following a diurnal cycle (Figure 4-2). By taking the integrand of the instantaneous fluxes over the year, the traditional positive NPP profile is seen, where the cropland areas amass organic material throughout the growing season (Figure 4-2).

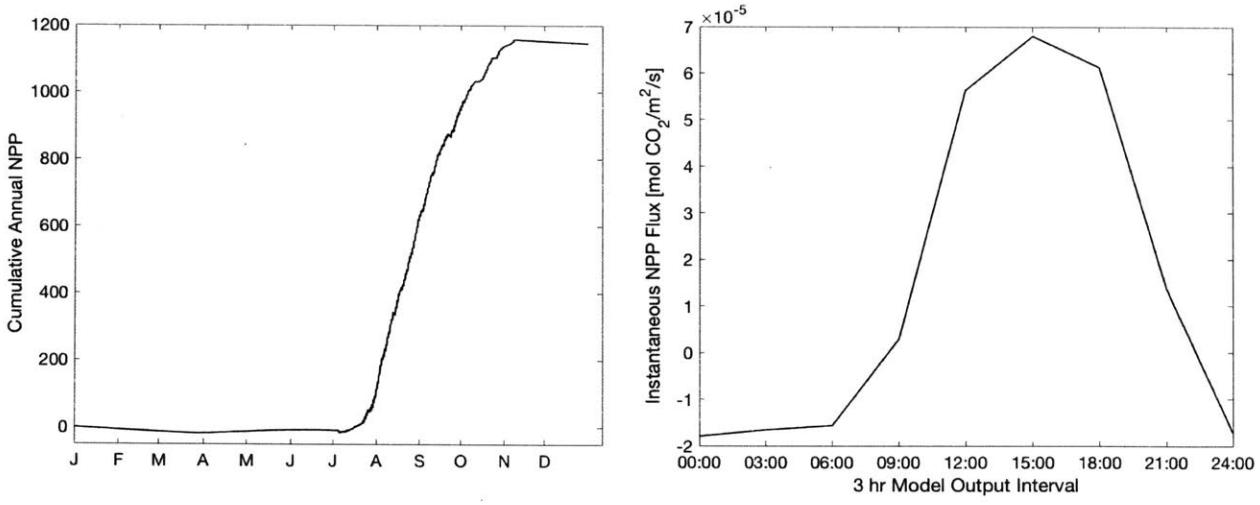


Figure 4-2: PDF of cumulative NPP for randomly selected cropland grid cell (43.78° N, 94.75° W) in 1986 (left). Diurnal cycle of NPP instantaneous flux for the same grid cell in mid-August (right)

4.2.3 NPP Data and Correlation to Crop Yield

While photosynthesis rate will be used in order to reproduce realistic rates of NPP within the model, it is important to understand how NPP relates to yield and can be used in order to represent agricultural productivity increases.

The advent of the satellite era has made high resolution datasets available for variables that were previously measurable only at individual sites, namely vegetative and soil characteristics. However, researchers must be careful not to make spurious correlations between indices and plant processes (Glenn et al., 2008). Oftentimes values are heavily dependent on input datasets, or assumptions about vegetation parameters and must be carefully calibrated from ground measurements. Glenn et al. (2008) notes that satellite based vegetation indices are best used for parameters associated with light uptake, and are less credible for deriving more involved canopy characteristics. One of the main problems associated with satellite derived cropland NPP in studies has been the assumption of constant Light Use Efficiency (LUE) across crop types, leading to underestimation of GPP and NPP in productive croplands (Bandaru et al., 2013; Jaafar & Ahmad, 2015; Li et al., 2014; Lobell et al., 2002; Reeves et al., 2005; Turner et al., 2006; Xin et al., 2015; Zhao et al., 2005). MODIS also has the problem of overestimating cropland area in some cases, causing an overestimation of total NPP, even when the mean is underestimated (Li et al., 2014).

Instead of a satellite based NPP estimate, this study will use the other main methodology for deriving net primary productivity — inventory based calculations. This method is beneficial to this particular study because it provides a direct physical link between crop

Table 4.1: Corn specific parameters for crop inventory yield to NPP conversion equation in the literature

Source	Mass / Unit Reported Yield	Moisture Content	Harvest Index	Aboveground Biomass Fraction	Carbon/Biomass Equivalence (gC/g)
Prince et al., 2001	25.401	13%	0.53	0.85	0.50
Lobell et al., 2002	25.401	11%	0.45	0.80/0.90	0.45
Hicke & Lobell, 2004	25.401	11%	0.45	0.85	0.45
West et al., 2010	25.4	13%	0.53	0.85	0.45
Li et al., 2014	25.4	13%	0.53	0.85	0.45

yield and NPP. The primary dataset and methodology described in this section is obtained from a study by Prince et al. (2001) which noted that crop inventories in high production areas can be used to reconstruct detailed temporal and spatial NPP datasets. Prince et al. (2001) looked at the NPP of corn, soybean, sorghum, sunflower, oats, barley, wheat and hay for areas in Illinois, Indiana, Iowa, Wisconsin, Michigan, Minnesota, North Dakota and Iowa for 1992. The dataset provides values for the centroid of 0.5° by 0.5° grid cells including total NPP, the percentage of cropland and forest cover for the grid cell, and the percentage of that cropland that was separately corn and soybean. The NPP of non-crop and non-woodland areas was assumed to be the same as the mean of the crop and wooded areas.

Prince et al. (2001) used agricultural inventory data from the USDA NASS in order to estimate total NPP estimates. The method consisted of a conversion of yield into plant mass. Prince et al. (2001) notes that their methodology is accurate within 50 $gC/m^2/yr$, and while Hicke and Lobell (2004) note that USDA data for the most productive crops can have a 95% confidence interval error of 5-10%, inventory statistics are widely accepted as high-quality measures of yield in the United States. While Prince et al. (2001) does not explicitly provide the equation, Hicke and Lobell (2004) uses the same methodology and dictates the equation as:

$$NPP\,[gC/m^2/yr] = \frac{\frac{Yield \times Mass\,per\,Unit\,Harvest \times (1 - Moisture\,Content) \times Carbon\,Equivalent}{Harvest\,Index \times Aboveground\,Fraction}}{Harvested\,Area}$$

This formulation has been used in several studies (Bandaru et al., 2013; Hicke & Lobell, 2004; Jaafar & Ahmad, 2015; Li et al., 2014; Monfreda et al., 2008; Prince et al., 2010; West et al., 2010). Importantly, there is consistency in the values of equation parameters across different studies (Table 4.1)

Prince et al. (2001) notes that there were two main potential sources of error in the estimations - root:shoot (R:S) ratio and Harvest Index (HI), factors which have also been isolated in other studies. R:S ratio relates to the partitioning of belowground and aboveground biomass and depends greatly on the cultivar. However, the ultimate NPP values in Prince et al. (2001) were not very sensitive to the exact R:S values. Harvest Index relates to the ratio of seed yield to aboveground biomass, and it is a parameter that has been shown to increase dramatically with crop yields, but not for corn (Long et al., 2006; Monfreda et al., 2008; Tollenaar & Lee, 2006). Tollenaar and Lee (2006) note that this is because both the dry matter accumulation and the number of kernels within corn have kept pace with each other and have maintained a constant HI even as corn yields have increased. Therefore, a consistent relationship can be assumed for corn between yield and NPP throughout the century based on the stability of these additional parameters. This consistency allows this study to describe a proportional relationship between yield and NPP that is implied through the formula and supported in the literature.

Hicke and Lobell (2004) examined agricultural productivity over the Central U.S. South Dakota, Nebraska, Kansas, Missouri, Iowa, Minnesota, Wisconsin, and Illinois) from 1972-2002 and found that NPP increased 46% (380 to 550 $gC/m^2/yr$) while production increased 51% in the same time-period. This corresponds to an NPP trend of 5.66 $gC/m^2/yr$. If it is assumed that this trend has been consistent since crop yields have begun to increase, a cropland NPP in this region in 1930 can be estimated as approximately 137 $gC/m^2/yr$. This corresponds to a 4 fold increase in NPP from 1930 to 2002. Correspondingly, using the weighted mean crop yield from all Midwestern states, a yield increase from a 1900-1935 average yield of 32.86 BU/acre in 1930 to 137.08 BU/Acre in 2002, represents a 4.17 fold increase in yield.

4.2.4 Conversion of Data

IBIS currently simulates cropland as being entirely composed of a C_4 crop such as corn, so the mean cropland NPP given by Prince et al. (2001) must be converted to corn only NPP to be consistent with the model setup. To do this the average woodland NPP contribution assumed in Prince et al. (2001) of 12 Mg/ha/yr with a carbon equivalent of 0.5 is subtracted from the total. It can also be assumed that corn NPP is on average about three times higher than that of other crops, particularly soybean. This is supported by plotting the proportion of yield between corn and soybean throughout the years (Figure 4-3), and also through values presented in Li et al. (2014) for corn and soybean NPP calculated from USDA NASS data. Li et al. (2014) estimated NPP to be 952 ± 163 $gC/m^2/yr$ for corn and 375 ± 74 $gC/m^2/yr$

for soybean in 2007, and 990 ± 141 gC/m^2/yr for corn and 352 ± 72 gC/m^2/yr for soybean in 2008.

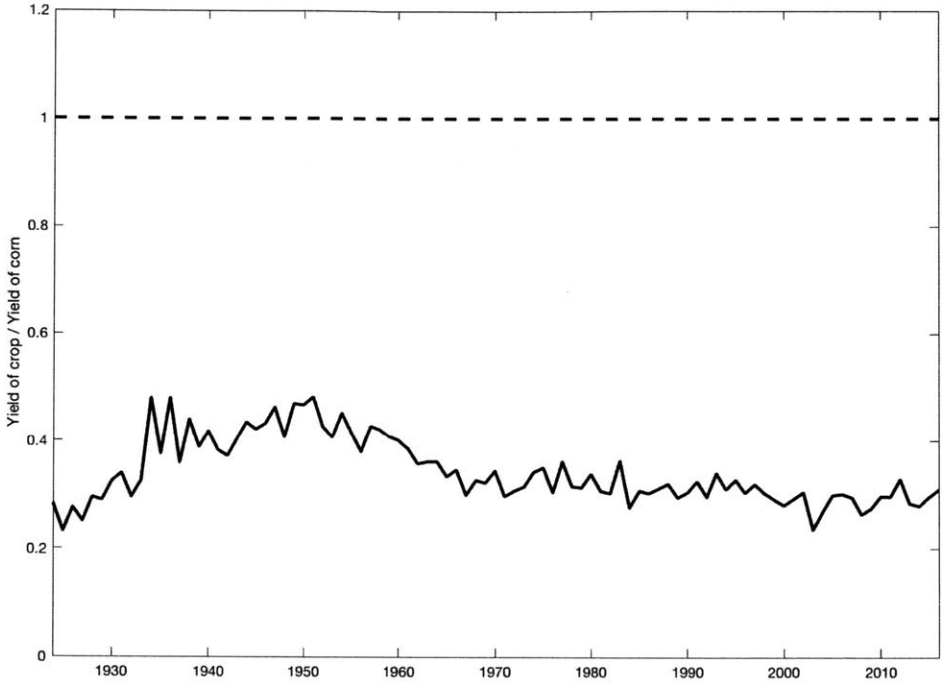


Figure 4-3: Comparison of corn yield (dotted line) and soybean yield (solid line) normalized by the corn yield in any given year. (Data Source: USDA-NASS, 2016)

The following formula is used to make the conversion:

$$Corn\ NPP\ [gC/m^2/yr] = \frac{NPP\ \times\ (crop\ area\ +\ forest\ area)\ -\ 600\ \times\ forest\ area}{corn\ area\ +\ other\ crop\ area\ \times\ \frac{1}{3}}$$

This makes the assumption that all crops included in the percentage given in the Prince et al. (2001) are either corn or soybean (even when the percentages of corn and soybean given do not equal 100%). This is an acceptable assumption because corn and soybean are the dominant crops in this region, and account for roughly 87% of the net primary productivity (Li et al., 2014). The final transformed Prince et al. (2001) data can be seen in Figure 4-4.

On average the corn NPP is about twice as large as the mean NPP in this region. This is supported by several figures from the literature. Li et al. (2014) lists the mean cropland NPP estimated from USDA data to be 660 ± 320 gC/m^2/yr in 2007 and 656 ± 330 gC/m^2/yr in 2008 while citing the USDA derived NPP of corn to be 952 ± 163 gC/m^2/yr and 990 ± 141

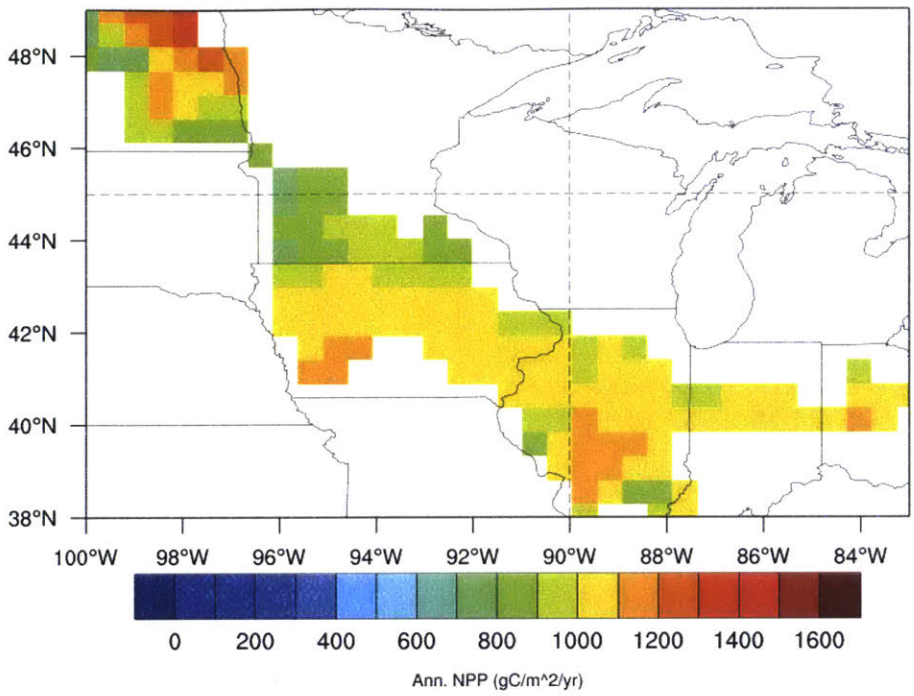


Figure 4-4: Corn equivalent annual NPP data for 1992 (Data Source: Prince et al., 2001)

$gC/m^2/yr$ for the same years, putting the corn NPP at double that of the mean within the range of errors of the estimates, although the paper itself cites corn NPP as being only 30% above the mean.

Hicke and Lobell (2004) give a mean cropland NPP in 2002 of 550 $gC/m^2/yr$. For the assumption of a 1100 $gC/m^2/yr$ average corn NPP in 2002, and a similar 4 to 5 fold increase since 1930, the corn NPP trend would need to be approximately 11.46 $gC/m^2/yr$ to 12.22 $gC/m^2/yr$. In a separate study, Twine and Kucharik (2009) found that corn in the central and eastern United States had a trend of 6.43 $gC/m^2/yr$ from 1982-2002 and a trend of 2.55 $gC/m^2/yr$ for 1950-2002. This study uses the more conservative trend as an estimate, as it spans a longer proportion of the 20th century. However, Twine and Kucharik (2009) were attempting to look only at climate impacts on yield, removing all other influence on NPP. They compared their NPP increases to that of Hicke and Lobell (2004) and found that their NPP increases likely represented only 20-25% of the actual increase. Using this it can be assumed therefore that the true NPP corn yield trend was approximately 10.2 to 12.75 $gC/m^2/yr$ for 1950-2002, matching what would be expected using an assumption of corn being twice as productive as the mean. This is of course dependent on the assumption that the trends from 1930-1950 matched those of the latter half of the century, but based on the linear increase in corn yields across this entire period, this is likely an appropriate

assumption.

It is important to note again the major assumption that is being made in matching the NPP magnitude with that of corn. Currently the model is simulating the entire area as if it was corn. This necessarily means that the amount of biomass produced by the model is higher than it would be if it simulated a realistic combination of crop type and other vegetation within cropland areas. However, this study chooses to maintain physical consistency with the productivity of corn.

4.2.5 Calibration Methodology

The calibration of the model depends entirely on the 1992 distribution of NPP from Prince et al. (2001) because it provides a corn based spatial estimate over most of the high productivity areas in the Midwest. Therefore, it is assumed that productivity of cropland in the Midwest as a whole is consistent on average with the areas that were quantified by the Prince et al. (2001) study. The goal of this calibration was first to determine a model multiplication factor for ancr that would reproduce observational NPP data, and then to use the established relationship between yield increases and NPP increases in order to determine how that factor would need to change throughout the century in order to accurately represent yield increases throughout time.

To this end, the same small scale two-year runs that were conducted in order to determine the model relationship between the multiplication factor and NPP output are again used here. For each of these, an average model NPP was calculated for each run using the average of all the irrigated and non-irrigated grid cells within the region 38.25 to 48.75° N and 83.25 to 99.75° W, which captured the extent of the Prince et al. (2001) data. The average 1992 cropland NPP produced by the model was 583 $gC/m^2/yr$ and 1128 $gC/m^2/yr$ for a multiplication factor of 1 and 2 respectively (See Figure 4-5). The average derived corn NPP from Prince et al. (2001) for 1992 was 1064 $gC/m^2/yr$. Using the linear relationship between multiplication factor and NPP output, it was determined that a multiplication factor of approximately 1.88 would be appropriate to represent corn NPP in 1992.

It is important to note that there is a restriction on plant growth designated within the model that will also affect annual NPP production. Growth within cropland areas is not allowed before July, and is forced within the model by setting the accumulated growing degree days for crops to zero before July. This helps restrict growth to the correct season and prevents high NPP too early in the year.

This scheduling is consistent with crop progress data released by the USDA NASS (USDA, 1999, 2016). For the 1994-1998 average, corn and soybean had emerged in significant numbers

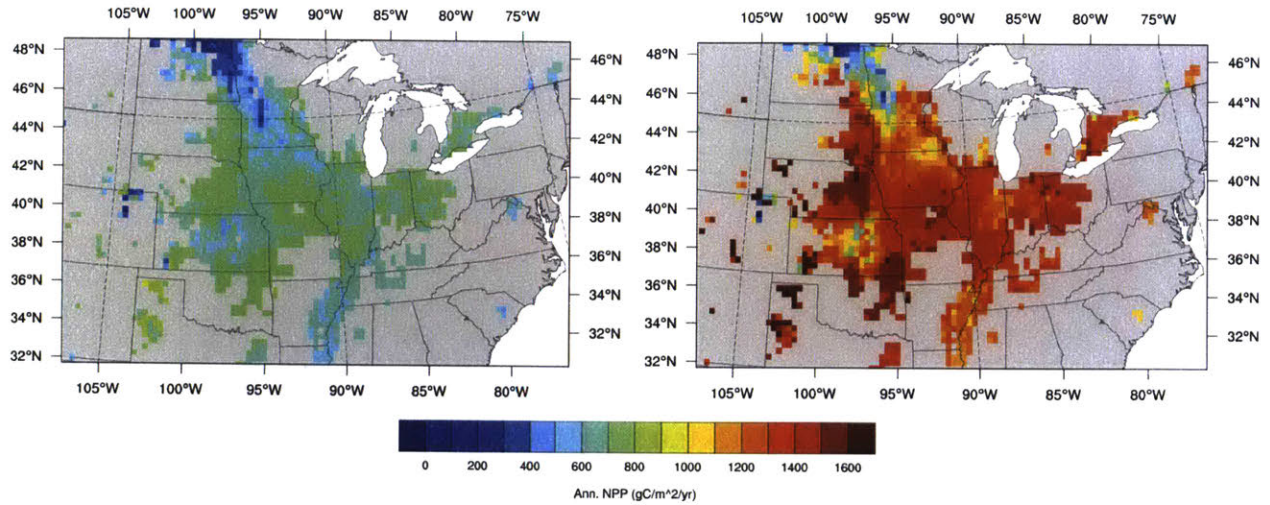


Figure 4-5: Model output of annual NPP ($gC/m^2/yr$) shown only for irrigated and non-irrigated grid cells. Model simulation with multiplication factor of 1 (left). Model simulation with multiplication factor of 2 (right).

by June but did not reach the mid-season silking and blooming phases in large percentages in any Midwestern states even by the beginning of July. For the 2011-2015 average, this is still broadly true although a higher percentage of corn and soybean had reached the silking and blooming phase by the first week of July, albeit with almost no states exceeding 25% of the total crop at this stage. For both time periods, the majority of the crop had been harvested by the first week in November and this is also consistent with the yearly cycle of NPP that the model produces (Figure 4-2 & Figure 4-6). In the future, it would be worthwhile to examine the impact of shifting and increasing this growth period, as studies have shown that even slight increases in growing season and changes in the length of crop development periods can have significant effects on yield and also on the energy balance effects associated with agriculture (Sacks & Kucharik, 2011).

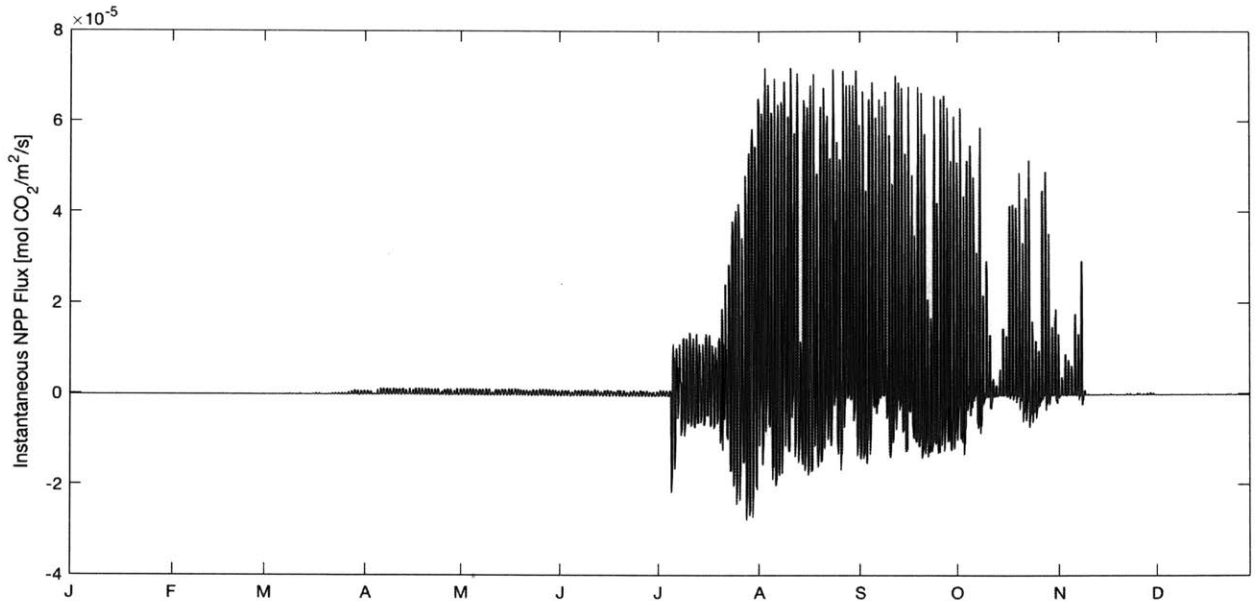


Figure 4-6: Plot of model output NPP instantaneous flux (mol-$CO_2/m^2/s$) for the same randomly selected grid-cell as in Fig. 4-2 (43.78° N, 94.75° W), for the year 1986

4.2.6 Determination of Parameter Series

Now that the correlation between yield increases and NPP increases over the past century has been shown, the well documented trend in corn yield can be used in order to determine the appropriate parameter series for NPP. The analysis is anchored in 1992 and uses the yield trend as an appropriate scale.

The weighted average Midwestern corn yield gives a yield of 125.65 BU/acre in 1992 (Figure 1-5). Based on the data, yields are essentially stagnant from 1900-1935, and a constant yield of roughly 32.86 BU/acre is appropriate for this period. The ratio of these two yields is 0.26 which will be used as the ratio applied to the multiplication factor between these two as well, since it has been shown that the increases in NPP scale roughly as the increases in yield. This gives a multiplication factor of 0.49 for the early period. Assuming, in accordance with the yield trend, that the changes between these two periods are linear, a time-series can be developed for the multiplication factors that will be applied to each decade of the model run (Table 4.2).

4.3 Model Validation

The Prince et al. (2001) dataset also provides a limited opportunity to determine whether the calculated multiplication factors were appropriate for creating a time-series of NPP, by

Table 4.2: Photosynthesis multiplication factors used in each deacde of model simulation in order to represent increases in crop yield.

Decade	Factor
1900	0.49
1910	0.49
1920	0.49
1930	0.49
1940	0.72
1950	0.95
1960	1.19
1970	1.42
1980	1.65
1990	1.88
2000	2.11

comparing to a dataset for the state of Iowa from 1982 to 1996. For this data, no information about the relative composition of the grid cells was given and so the average conversion of corn $NPP = mean\ NPP \times 2$ from the previous data was applied to this Iowa subset of the Prince et al. (2001) data. This will likely produce higher "corn" NPP values for Iowa because the state already has both the highest acreage of corn harvested and the highest production of all Midwestern states (Appendix C). In order to compare to this Iowa time-series, a simulation was run from 1982-1995 and split into two segments. A January 1982 to December 1985 simulation was run using the 1980 land use map, and a multiplication factor of 1.65 to match the developed time-series. This was combined with a January 1986 to December 1995 simulation run with the 1990 land use map and a multiplication factor of 1.88. Results of average annual NPP for Iowa from the data and model are shown in Figure 4-7. Cropland NPP was averaged over a region from 40.75 to 43.25°N and 90.75 to 96.5°W to approximate the Prince et al. (2001) Iowa region.

The 1982-1985 decadal average NPP for the model in this region is 903 $gC/m^2/yr$ while the Prince et al. (2001) data yielded 997 $gC/m^2/yr$ — a difference of 94 $gC/m^2/yr$. The 1986-1995 decadal average of NPP for the model was 1017 $gC/m^2/yr$ while the Prince et al. (2001) data yielded $gC/m^2/yr$ — a difference of 54 $gC/m^2/yr$. Note that there are several periods where the model and the inventory based NPP data are severely mismatched. This occurs in 1988 and 1993 in correspondence to a major drought and flood year respectively. Inventory based methods are erroneous when quantifying annual NPP in years where major events decimate crop yields, because they are based on only the biomass harvested, and not

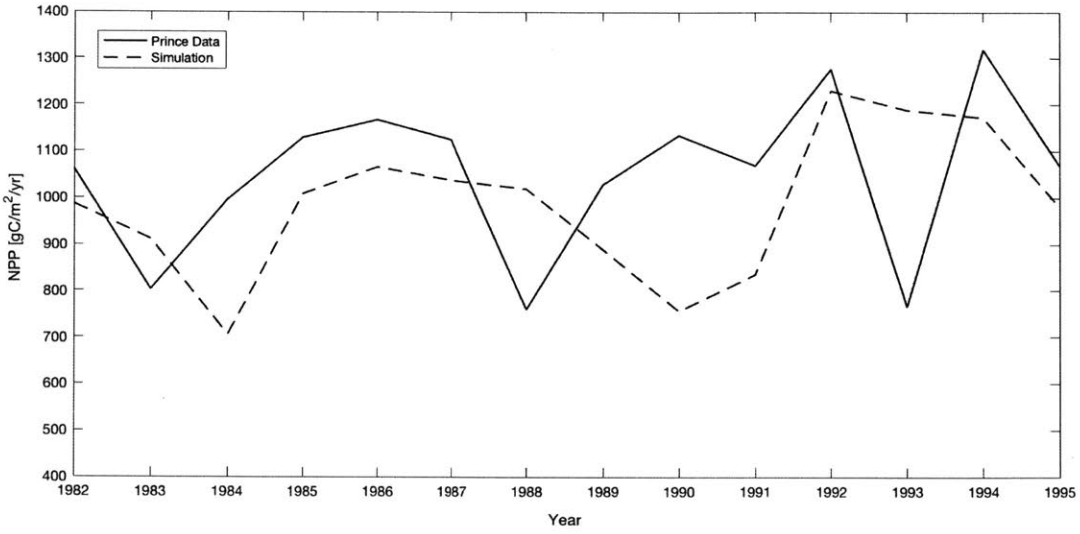


Figure 4-7: Comparison of results for Iowa averaged NPP from Prince et al. (2001) corn equivalent (solid line) and model simulation output (dotted line)

the growth during the full season. The misstep between the model and the Prince et al. (2001) data in 1990 is not attributed to any major weather event.

The appropriateness of a spatially uniform application of a multiplication factor must also be considered. This implies that the increases in NPP have been constant throughout the majority of the Midwest. It is likely a valid assumption that management practices and crop breeds have been uniformly adopted throughout the United States over the 20th century with slight differences in individual farms. Various maps of NPP trends for the latter half of the century seem to support the assumption of uniformity (Hicke et al, 2004, Fig. 6a; Twine & Kucharik, 2009, Fig. 3a-b; Mueller et al., 2016, Fig. 3e).

The NPP trend shown in the references figures is spatially uniform throughout most of the Midwest, especially relative to regions in the West and Southeast. However, this trend is both steeper and more spatially heterogeneous when looking at only the last few decades versus the last half century (Twine & Kucharik, 2009, Figure 3a,b). Because the current study attempts to fit the NPP trend for the full century, it uses the trends and spatial homogeneity attached to the long term 1950-2002 average (Twine & Kucharik, 2009, Figure 3a). Also important is the evidence that irrigated areas have seen a dramatically higher increase in yield than non-irrigated cropland, and a reduction in interannual yield variability (Kucharik & Ramankutty, 2005). This is something that should be taken account in the future when trying to spatially delineate yield increases, but for this study, the inclusion of a single crop type in the model and the small spatial extent of irrigation relative to non-irrigated cropland within the domain mean that this difference was not taken into account

explicitly.

4.3.1 Temperature and Precipitation Comparison

First, validation that the photosynthesis multiplication factor, NPP, and yield all maintain their relationship within the model throughout the full century and the full range of multiplication factors applied is shown in Figure 4-8. NPP sees a similar scale of increase relative to yield as expected, and while it does not precisely match the interannual variability of the yield trend it does show a similar increase in the year to year interannual variability in the latter half of the century.

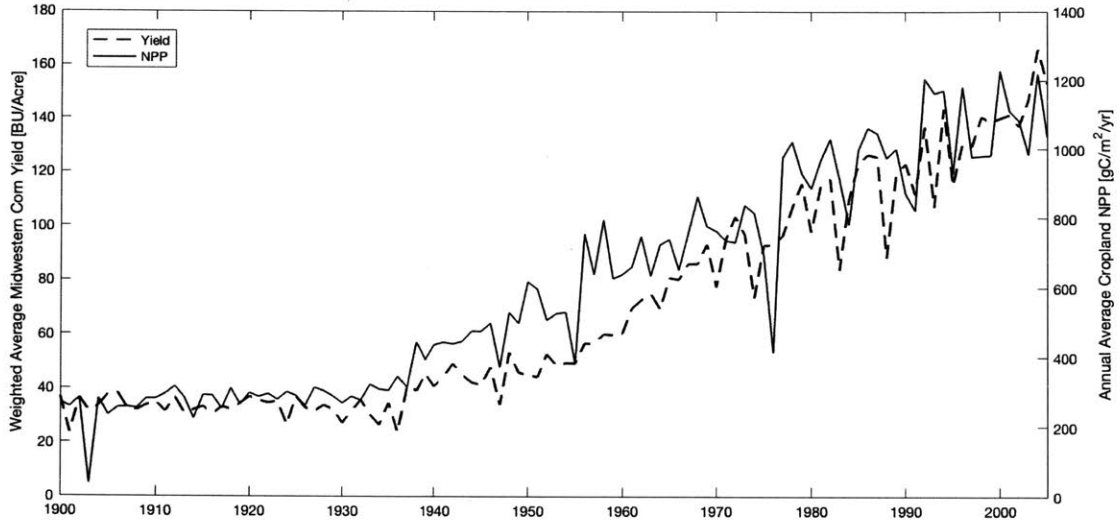


Figure 4-8: Weighted average Midwestern corn yield (dotted line) plotted with average annual cropland NPP from ERA20C VG simulation (solid line)

Before running the full model simulations with this multiplication factor time-series a run from 1982-1996 was compared to temperature and precipitation observations to ensure that the change of the photosynthesis parameter was not causing changes in temperature and precipitation that were unrealistic relative to the model default. It is important to note here that this 1982-1996 simulation was run with the original set of boundary conditions — ERA20C. While this should not affect the results of the NPP validation and the selection of photosynthesis parameter based on the internal model relationship of those two variables, it will have an impact on the temperature and precipitation results shown here. While the ERA20C reanalysis surface temperature follows the general pattern of temperatures that are present in the gridded CRU and UDEL datasets, it has a pronounced June-August cold bias in the latter part of the 20th century in the study region and this will bias the model output.

The ROSC averaged surface temperatures for June-August are plotted below for UDEL, CRU and the model output as well as the ERA20C reanalysis 2m surface temperature (Figure 4-9). The same comparison is performed for average precipitation over the ROSC, however in this case there is no reanalysis data to compare to (Figure 4-10). The differences are tabulated as appropriate decadal averages in Table 4.3 & 4.4. Using the series of multiplication factors outlined in this chapter, this study can move forward with a representation of the large crop intensification that has occurred in the Midwest over the 20th century.

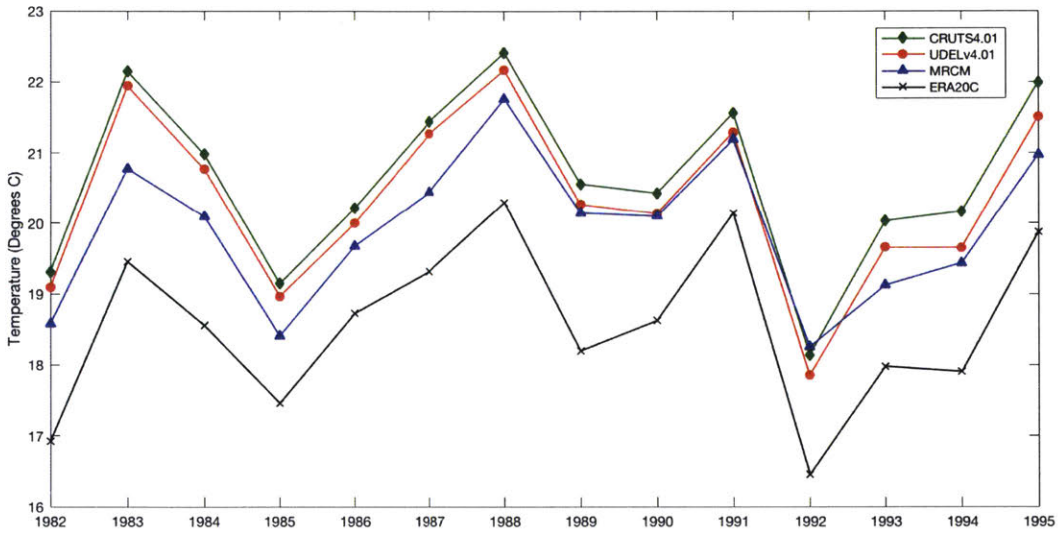


Figure 4-9: Comparison of JJA ROSC averaged temperatures from 1982-1996. Comparisons are between CRUTS4.01 (green), UDELv4.01 (red), the ERA20C VG simulation results (blue) and the ERA20C reanalysis 2m surface temperature (black)

Table 4.3: JJA Temperature Comparison (° C)

	MRCM	CRU	UDEL
1982-1985	19.46	20.39	20.19
1986-1995	20.11	20.69	20.38

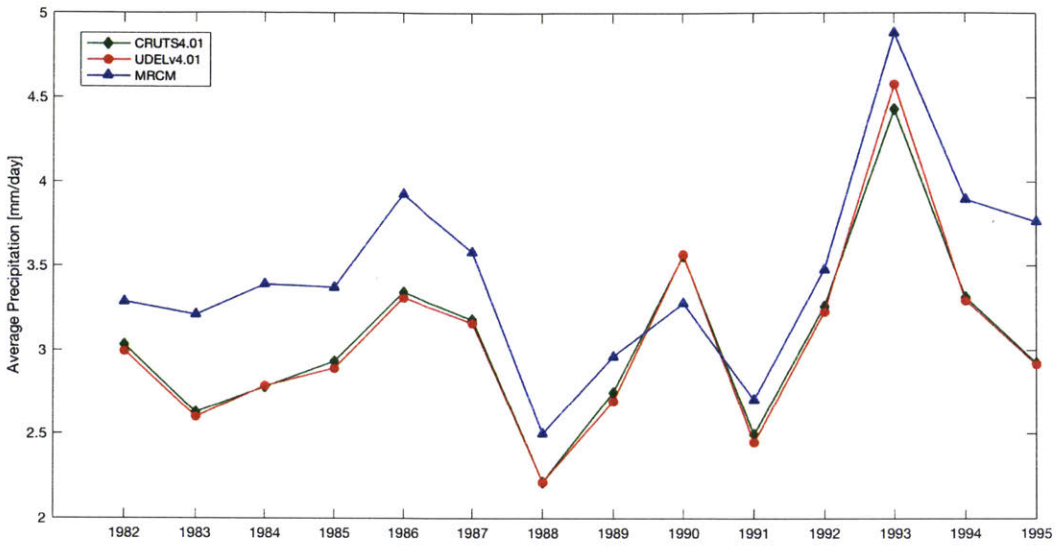


Figure 4-10: Comparison of JJA ROSC averaged daily precipitation(mm/day) from 1982-1996. Comparisons are between CRUTS4.01 (green), UDELv4.01 (red), and the ERA20C VG simulation results (blue)

Table 4.4: JJA Precipitation Comparison (mm/day)

	MRCM	CRU	UDEL
1982-1985	3.31	2.84	2.82
1986-1995	3.50	3.15	3.14

Chapter 5

Simulation Results and Analysis

The results of two full sets of three simulations each are presented here, with the only difference between the two being the reanalysis data used as lateral boundary conditions – either ERA20C or CERA20C. All simulations are run on a domain centered at 40.5° N and 91.5° W with a 30-km grid spacing, and 122 zonal points and 80 meridional points. This corresponds to a domain of roughly 30.4 – 50.8° N, and 112 – 70.6° W. The ERA20C simulations are run from January 1, 1900 through December 31, 2005 and the CERA20C simulations are run from January 1, 1901 through December 31, 2005.

Each of the two sets contains the same three simulations:

1. *No Vegetation Development, No Greenhouse Gas Increases* (**NVNG**): This simulation uses the 1900 land use map and maintains 1900 greenhouse gas levels along with a photosynthesis multiplication factor of 0.49 for the entirety of the run.

2. *No Vegetation Development, Realistic Greenhouse Gas Series* (**NVG**): This simulation also uses the 1900 land use map and a constant photosynthesis multiplication factor of 0.49, but greenhouse gases evolve throughout the century, consistent with the CMIP5 historical time-series shown in Chapter 3.

3. *Realistic Vegetation Development, Realistic Greenhouse Gas Series* (**VG**): This simulation is run in decadal segments (except the first simulation which is run from 1900 or 1901 to 1905) with the land use map being updated to reflect cropland and irrigation expansion in the 10 years surrounding 1900, 1910, 1920, etc. For instance, the simulation years connected to the 1920 land use map are 1916-1925. At each of these points the photosynthesis multiplication factor is also changed to reflect the value developed for the decade as discussed in Chapter 4. Again, greenhouse gases evolve throughout the century, consistent with the CMIP5 historical time-series.

For the VG simulation in each of the sets, the model was run with one extra year at the beginning of each of the decadal simulations starting January 1, and was not used in the post-processing and the construction of the full time-series. Soil moisture conditions are equilibrated for the region with long term offline simulations of IBIS, and no longer spin up time was deemed necessary. The model was not run continuously for VG due to the nature of the model save files generated, the way that they are utilized for a restart, and the current pre-processing procedure and setup for surface boundary conditions.

Each of the sets will first be compared to observational data, for the region broadly and in a smaller region (37 – 42.5° N, 87.5 – 91.5° W) to the Illinois State Water Survey. Then, they will be decomposed based on the individual impacts of greenhouse gases and agricultural development. This will also include a discussion of the influence that the boundary conditions themselves had on the results.

The analysis will be performed for two main regions:

1. **Domain**: this will refer to the average of the domain less the nine outer grid cells to eliminate any edge effects that may occur due to the boundary conditions. This makes the domain average correspond to output coordinates of 32.85 – 48.42° N and 108.92–73.73° W.

2. **Region of Significant Change (ROSC)**: this region from 39 – 48° N and 100 – 82° W, as identified by Douglas (2016), is a region in the Midwest that has seen significant observed changes in temperature and precipitation.

The time periods of comparison will be 1920-1949 (early period) and 1970-1999 (late period) to fit into the simulation time-period, again consistent with Douglas (2016). Additionally, all averages are shown for July-August. This time period is chosen to capture the time-period of maximum vegetative growth and irrigation impacts (DeAngelis et al., 2010; Ozdogan et al., 2010), and to allow for comparison to previous studies.

Where appropriate, area averaged changes are accompanied by the results of Kolmogorov-Smirnov two sample test (K-S test). This test determines whether two samples originate from the same distribution, but is non-parametric i.e. makes no assumption about the distribution of the sample itself (Sheskin, 2007). In a single sample K-S test the empirical distribution is compared to a cumulative distribution function, with the test statistic being dependent on the vertical distance between the two distributions. In a two sample test this same procedure is followed, but using two empirical distribution functions instead (Heckert & Filliben, 2003; Sheskin, 2007). This test is ideal for characterizing climate data as it makes no assumptions about distribution. Notably, the K-S test with two independent samples is

also more conservative — less likely to reject the null hypothesis — and therefore will tend to under-identify significantly different distributions relative to other similar tests such as Mann-Whitney U Test or the t test for two independent samples (Sheskin, 2007). Significance here is determined with a 95% confidence interval, i.e. a change is significant when the test yields a p value of less than 0.05 and the data likely come from two different distributions. In each case where a significance test was performed, the resulting p value is included in parentheses.

5.1 Comparison to Observations

For the observational analysis, all comparisons will be made relative to the VG simulations from either set. These simulations should contain all relevant forcings, and therefore the results should be the most representative of the observed climate. Based on available data, observational analysis will be made here for temperature, precipitation, specific humidity, and various radiation parameters. Further figures for CERA20C VG and ERA20C VG results for evapotranspiration, sensible heat, surface soil moisture, and root zone soil moisture can be seen in the appendix (B-18 – B-21).

There are two long-term observational datasets with good coverage and accuracy over the United States for the full 20th century. The first is the University of Delaware Precipitation and Temperature Time Series (v4.01) which spans 1900-2014 at 0.5° by 0.5° resolution (Willmott & Matsuura, 2015a, 2015b). While this data set has been rigorously tested, like other historical reconstructions, it suffers from reduced accuracy in time periods with fewer stations. The documentation notes that over the reconstruction time period, the number of stations contributing to monthly temperature values ranges from 3,000 to 18,000 while for precipitation this number is 7,500 to 40,500. A main source of data are the Global Historical Climate Network (GHCN) stations, which NOAA notes reached maximum density for measurements of temperature and precipitation in the 1960s (Peterson & Vose, 1997). GHCN3 was used for temperature reconstruction along with several other supplementary sources, while GHCN2 was used for precipitation. The second observational dataset used here is the Climatic Research Unit Time Series (v4.01) developed by the University of East Anglia which spans 1901-2016 in 0.5° by 0.5° resolution (Harris et al., 2014).

For each of these grid cells, in either data set, there one monthly mean value of total precipitation and surface temperature. While these datasets are coarser resolution than the model output they are averaged over a similar area for comparison. CRU and UDEL do not provide data over the entirety of the domain, and therefore ocean areas where values are missing are not included in the associated averages.

5.1.1 Temperature

The ROSC average temperatures for 1920-1949 JA for CRU, UDEL, ERA20C, and CERA20C are 21.51, 21.48, 21.50, and 22.78 °C respectively. This is a difference from the average of the observations of +0.1 and +1.29 °C for ERA20C and CERA20C. For the domain, the corresponding averages are 22.62, 22.52, 22.18, and 23.31 °C. This is a difference from the average of the observations of -0.39 and +0.74 °C for ERA20C and CERA20C. As can be seen in Figure 5-1, the CERA20C simulation matches the temperature distribution in the observations more closely, particularly in the southern portion of the domain, however it is still hotter that the observational data over both regions while the ERA20C simulation matches this time period much more closely, especially in the ROSC.

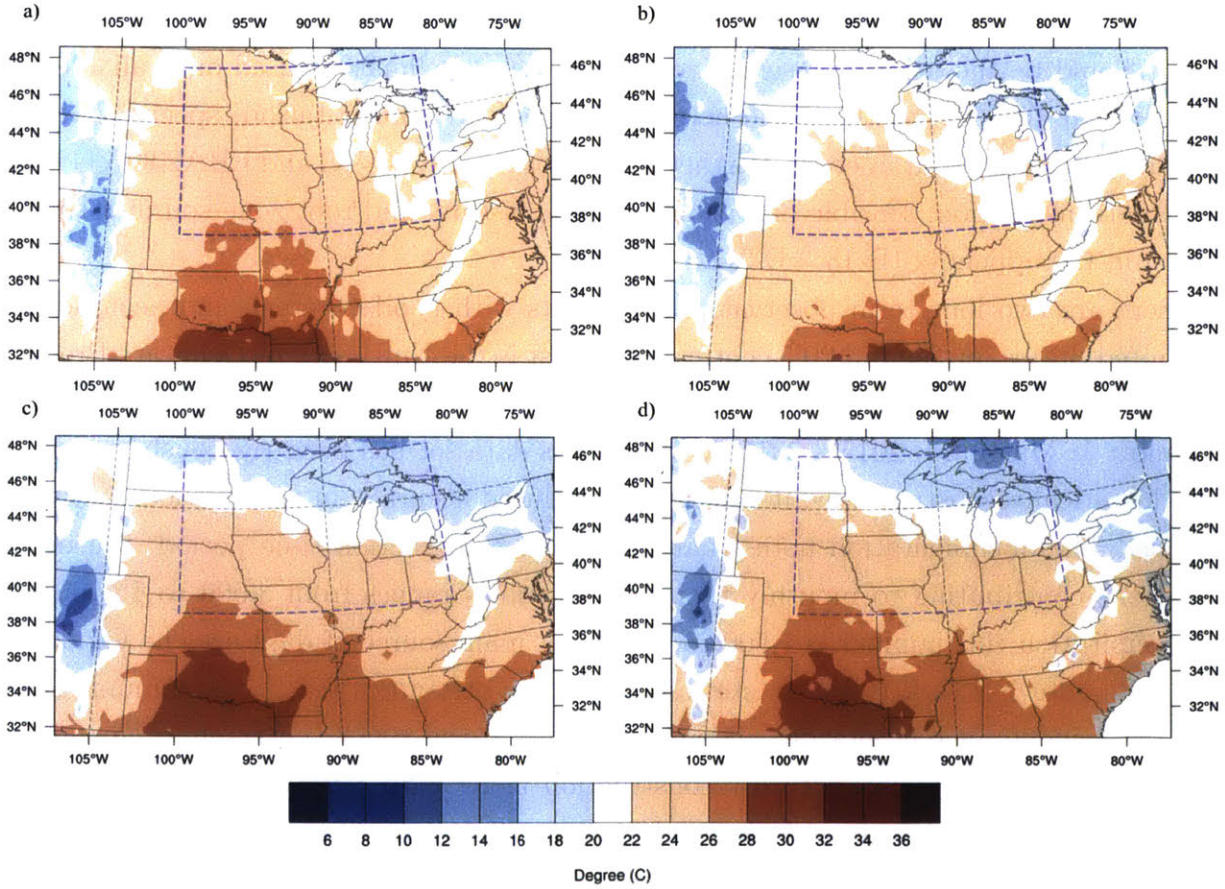


Figure 5-1: July-August 1920-1949 average surface temperature (°C) maps from a) CERA20C VG b) ERA20C VG c) CRUTS4.01 d) UDELv4.01. The dotted blue box denotes the ROSC.

Figure 5-2 shows the same comparison for the 1970-1999 average in each of the four datasets. The ROSC average temperatures for 1970-1999 for CRU, UDEL, ERA20C, and CERA20C are 21.21, 21.06, 20.31, and 21.91 °C respectively. This is a difference from the

average of the observations of -0.83 and +0.78 °C for ERA20C and CERA20C. For the domain, the corresponding averages are 22.45, 22.19, 21.52, and 22.93 °C This is a difference from the average of the observations of -0.8 and +0.61 °C for ERA20C and CERA20C.

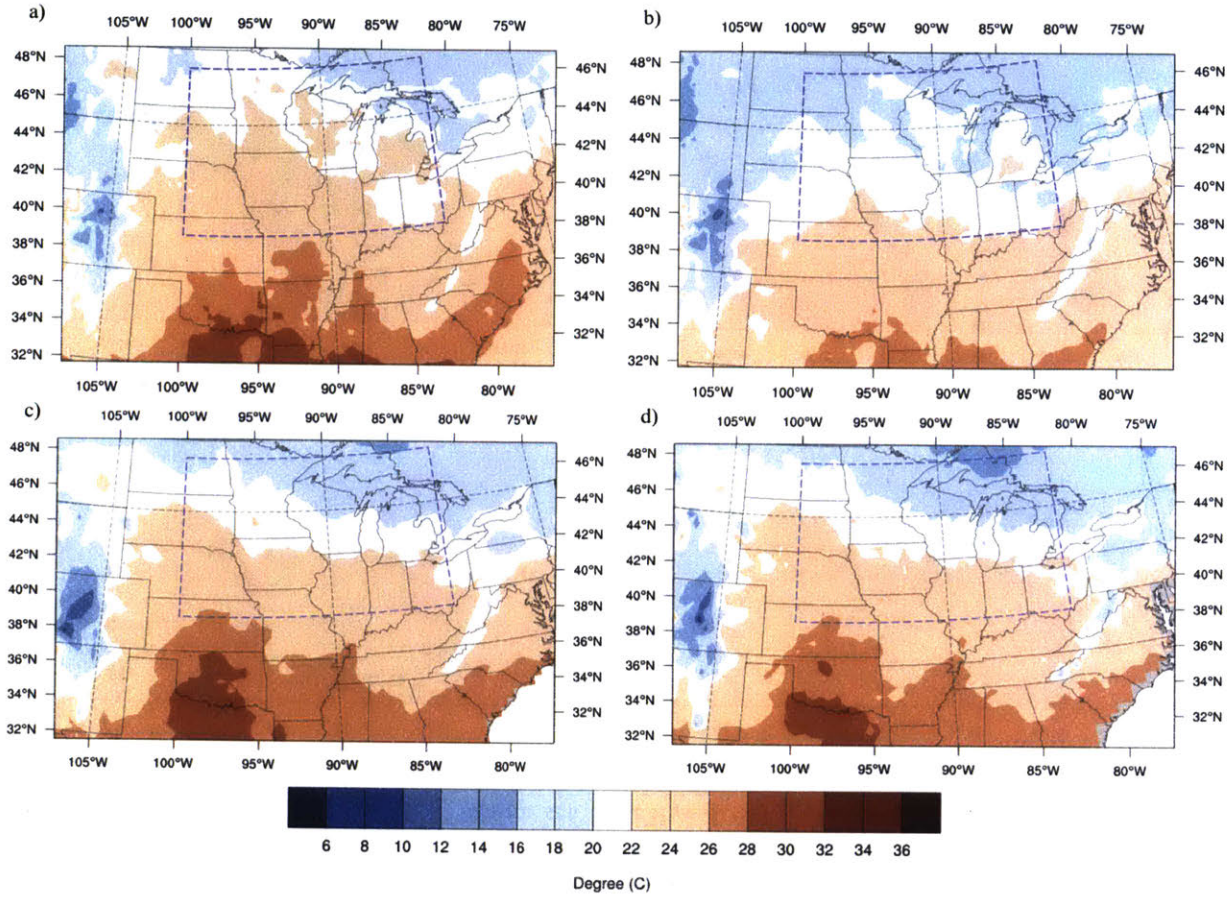


Figure 5-2: July-August 1970-1999 average surface temperature (°C) maps from a) CERA20C VG b) ERA20C VG c) CRUTS4.01 d) UDELv4.01. The dotted blue box denotes the ROSC.

The changes in average surface temperature within each grid cell from 1920-1949 to 1970-1999 are shown in Figure 5-3. The changes in ROSC average temperatures between the early and late periods for CRU, UDEL, ERA20C, and CERA20C are -0.29, -0.42, -1.18, and -0.87 °C respectively, with negative values denoting a decrease in average temperature from the early to late period. For the domain, the corresponding averages are -0.18, -0.32, -0.66, and -0.37 °C. CERA20C therefore matches the magnitude of temperature change over the domain over the two periods, but as can be seen in Figure 5-3. However, this is due to a balance between an overestimated cooling in the north and north west and an overestimated warming in the southeast relative to the observations. The temperature changes were measured as significant in ERA20C and CERA20C for both the ROSC ($p = 0.000017$ and $p = 0.0259$) and the domain averages ($p = 0.000017$ and $p = 0.0046$). Changes over the same time periods

were not considered significant in the CRU and UDEL data in the ROSC (p = 0.3420 and p = 0.2003) or the domain (p = 0.5372 and p = 0.2003).

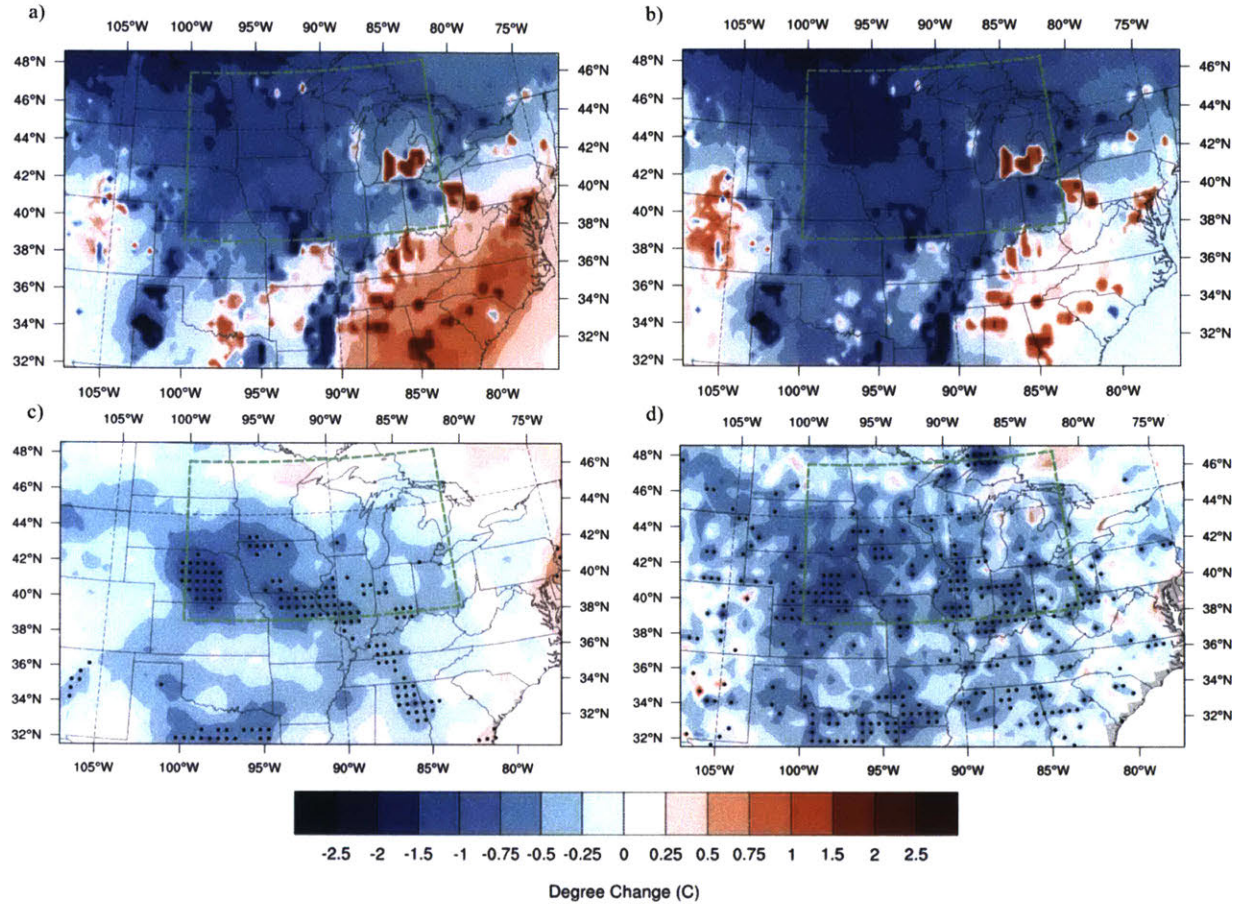


Figure 5-3: July-August temperature change (°C) from 1920-1949 to 1970-1999 in a) CERA20C VG b) ERA20C VG c) CRUTS4.01 d) UDELv4.01. The dotted green box denotes the ROSC. Black dots on c) and d) indicate grid cells with significant change at the 5% level with a K-S Test.

While both of the simulations overestimate the magnitude of cooling in the ROSC and domain from the early to late period, the patterns in Figure 5-3 describe how the simulations are performing and where any failures may stem from. The model matches the shape of cooling regions fairly well, especially when comparing the left portion of the domain. Cooling seems to occur throughout the full north-south span of the Midwest and Great Plains, with a weaker signal and even some warming in Colorado and New Mexico. The broad features of the differences between the model and the observations are the strong cooling signal that appears in the northwest of the domain, and the strong heating that appears in the southeast, particularly in the CERA20C simulation. This heating appears in the CRU data and the cooling appears to a greater degree in the UDEL data but neither simulations approach the 1.5 to 2.0 degree swing to either direction that the model results exhibit.

The pattern of cooling in the CRU and UDEL data is concentrated mostly in the Midwest, with a large portion of the grid cells exhibiting significant change falling within the ROSC. This is consistent with the idea that agriculture, and particularly agricultural intensification, results in cooling. These areas also show cooling in the model although to a stronger degree than in the observations. The observational data in both cases also shows a strong cooling signature in eastern Nebraska which corresponds to a large swath of irrigation that develops there beginning in the 1950s. There seems to also be a signature of cooling in the observational data along the entire irrigation corridor from the Texas panhandle to Nebraska, although the intense spots of cooling in Texas and Kansas only appear in the model results.

CERA20C and ERA20C are similar regarding areas of cooling but there are several differences that can likely be attributed to the background conditions imposed by the various reanalysis data sets as will be explored in a later section. ERA20C shows weaker warming in the southeast, but a stronger warming in the west over Colorado as well as a much stronger cooling in the southern Great Plains relative to the mix of warming and cooling in this area in CERA20C. Additionally, ERA20C shows a stronger anomalous cooling signature in the northwest portion of the domain than CERA20C.

5.1.2 Precipitation

Precipitation is analyzed in the same way with Figure 5-4 showing the July-August average distribution of daily precipitation for 1920-1949. The ROSC average daily precipitation for this early period for CRU, UDEL, ERA20C, and CERA20C is 2.49, 2.52, 3.06, and 2.79 mm/day respectively. This is a difference from the average of the observations of +0.56 mm/day (22%) and +0.29 mm/day (11%) for ERA20C and CERA20C. For the domain, the corresponding averages are 2.51, 2.56, 2.89, and 2.67 mm/day. This is a difference from the average of the observations of +0.36 mm/day (14%) and +0.14 mm/day (5%). In both regions, the CERA20C simulation matches the observations more closely. Both simulations capture the high precipitation along the east coast and in the Appalachians. The simulations also capture the corridor of higher relative precipitation in the center of the ROSC (Wisconsin, Illinois, Iowa), however the precipitation rate is overestimated in both simulations.

The ROSC average daily precipitation for 1970-1999 (Figure 5-5) for CRU, UDEL, ERA20C, and CERA20C is 2.96, 2.92, 3.11, and 2.88 mm/day respectively. This is a difference from the average of the observations of +0.17 mm/day (6%) and -0.06 mm/day (2%) for ERA20C and CERA20C. For the domain, the corresponding averages are 2.71, 2.72,

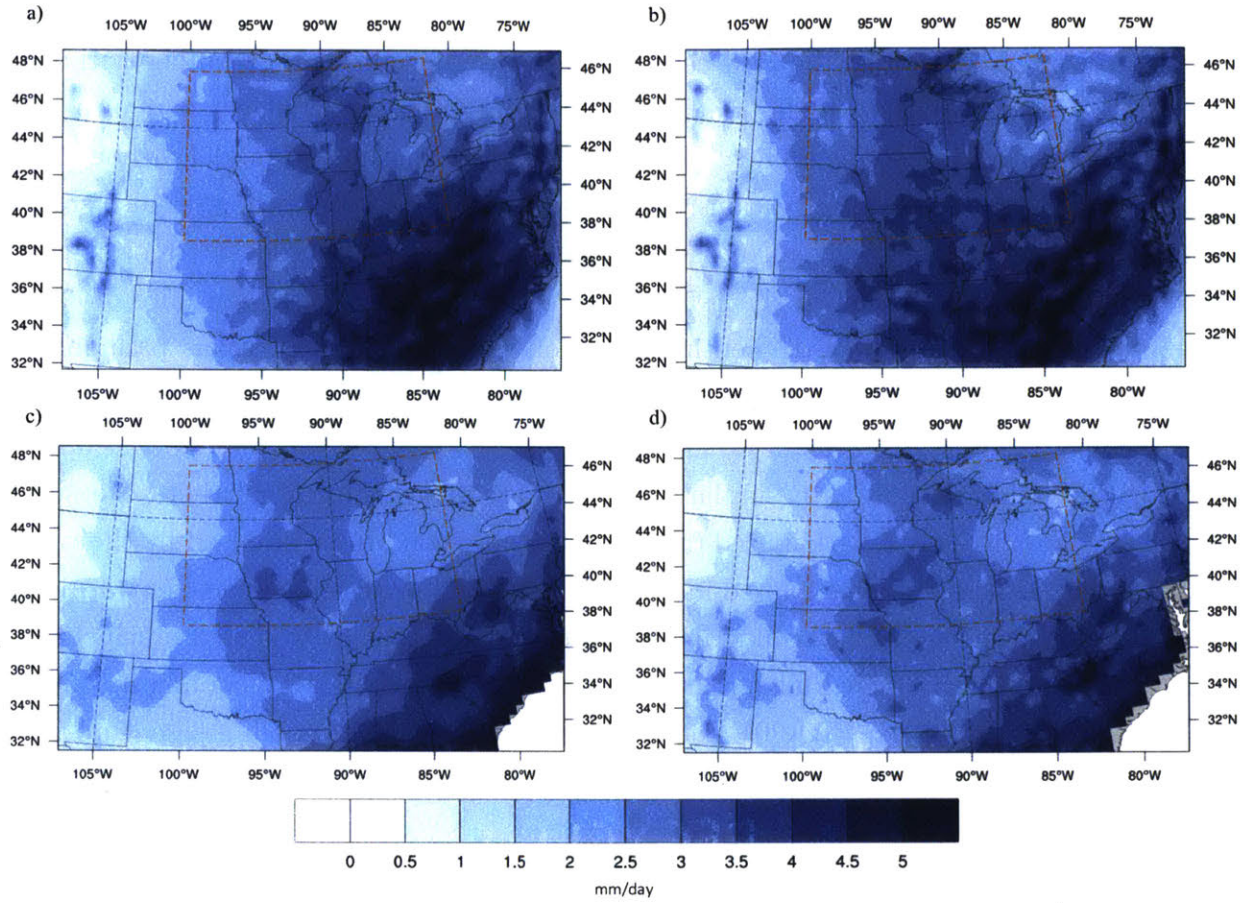


Figure 5-4: July-August 1920-1949 average daily precipitation (mm/day) maps from a) CERA20C VG b) ERA20C VG c) CRUTS4.01 d) UDELv4.01. The dotted red box denotes the ROSC.

2.89, and 2.59 mm/day. This is a difference from the average of the observations of +0.18 mm/day (8%) and -0.13 mm/day (5%). In both regions, the CERA20C simulation again matches the observations more closely, although the averages of both simulations are closer to the observations in both cases. Both simulations again capture the high precipitation along the east coast and in the Appalachians as well as the high precipitation ROSC strip, and the magnitude of rainfall is more accurately captured in these regions. The highest precipitation rates in the observational data are found directly along the east coast, while this is pushed inland in the model.

The changes in average daily precipitation within each grid cell from 1920-1949 to 1970-1999 are shown in Figure 5-6. The changes in the ROSC between the early and late periods for CRU, UDEL, ERA20C, and CERA20C are 0.47, 0.40, 0.06, and 0.09 mm/day respectively, with positive values denoting an increase in average precipitation from the early to late period. For the domain, the corresponding averages are 0.19, 0.16, 0.0052, and -0.086 mm/day, less than half of the corresponding ROSC values. In all cases both the observations

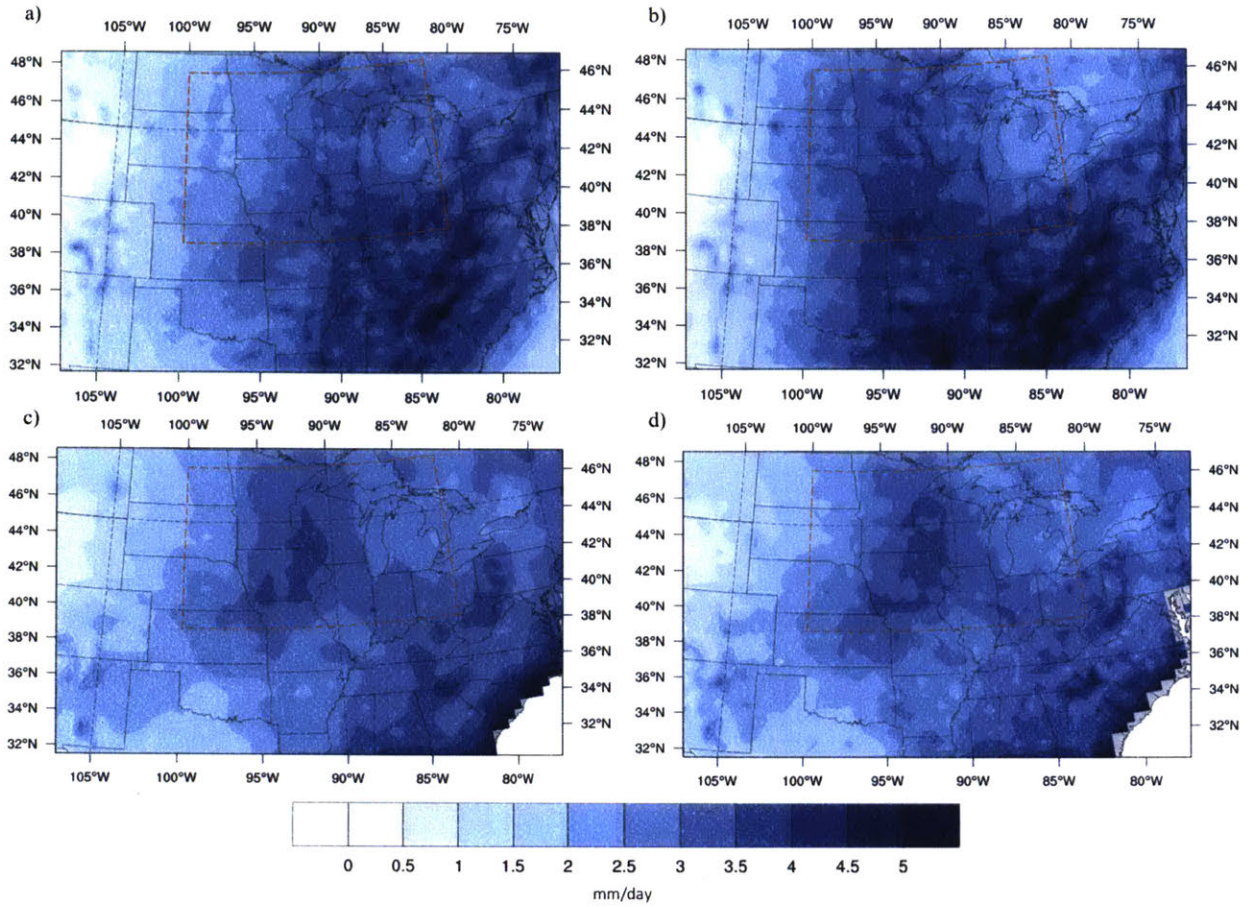


Figure 5-5: July-August 1970-1999 average daily precipitation (mm/day) maps from a) CERA20C VG b) ERA20C VG c) CRUTS4.01 d) UDELv4.01. The dotted red box denotes the ROSC.

and model simulations show an increase in precipitation over the ROSC, although the VG simulations for each reanalysis set vastly underestimate the magnitude of this change. Over the domain, the CERA20C simulation show a decrease in average precipitation albeit by a small amount. The precipitation changes were measured as significant in CRU and UDEL in the ROSC ($p = 0.0046$ and $p = 0.0113$) but not significant in ERA20C and CERA20C ($p = 0.2003$ and 0.1088). Over the same period, precipitation increases were not found to be significant in any of the four datasets in the domain, although CRU was significant at slightly below the 95% confidence level ($p = 0.0550$).

In similar terms to temperature, general patterns of change are examined to see how the simulation is reproducing historical observed change. In this case the dominant pattern of dry-wet-dry moving from the northwest to the southeast corner is captured in the simulations, though more clearly in the CERA20C simulation. The CERA20C simulation also captures the observed wetting in the northeast corner of the domain, a signal that is completely reversed in the ERA20C. The drying in Colorado and Wyoming is present in all four sets

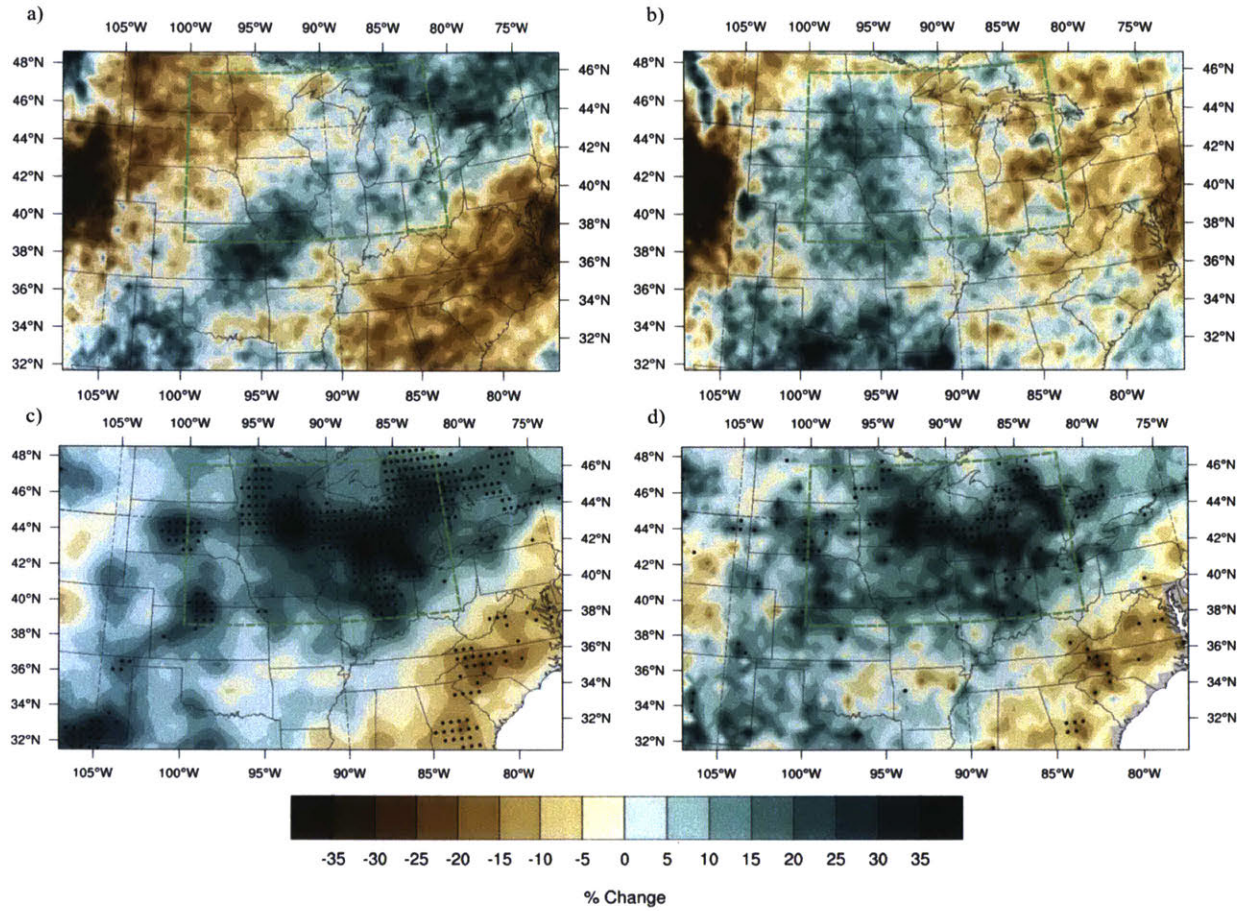


Figure 5-6: July-August daily average precipitation change (%) from 1920-1949 to 1970-1999 in a) CERA20C VG b) ERA20C VG c) CRUTS4.01 d) UDELv4.01. The dotted green box denotes the ROSC. Black dots on c) and d) indicate grid cells with significant change at the 5% level with a K-S Test.

of data as is the tendency towards wetting in the southwestern portion of the domain. The strongest precipitation feature in the observations — wetting in the ROSC — however, is not present as a distinct feature in either of the simulations though both show an average increase in precipitation in this area. Like with temperature, most of the strong wetting trends in the observations occur in areas of high agricultural productivity growth. There does not seem to be any spatial correlation with pockets of irrigation in TX, KS, or NE.

While precipitation changes are presented here in percentage terms, it is useful when looking at a wide region to acknowledge the absolute changes, since regions facing the same absolute changes in precipitation may seem to be more strongly effected by the nature of very wet or very dry base conditions. Precipitation change in absolute amounts can be seen in the Appendix (B-17). Some features such as the intense drying in the west are tempered by an absolute view due to the aridity of the area, while drying in the southeast is more

prominent.

5.1.3 Specific Humidity

Due to the lack of gridded observations containing specific humidity data, the NOAA 20th Century reanalysis dataset (NOAA 20CR V2c) is used for specific humidity data spanning 1851-2014 at 2° by 2° resolution. This reanalysis dataset is described in more detail in Chapter 3. Note that these are not historical observations, but the result of the assimilation of observations into a global climate model. Because these are not strictly observations, only the change between the early and late period is looked at in order to see the shape of the general trend (Figure 5-7).

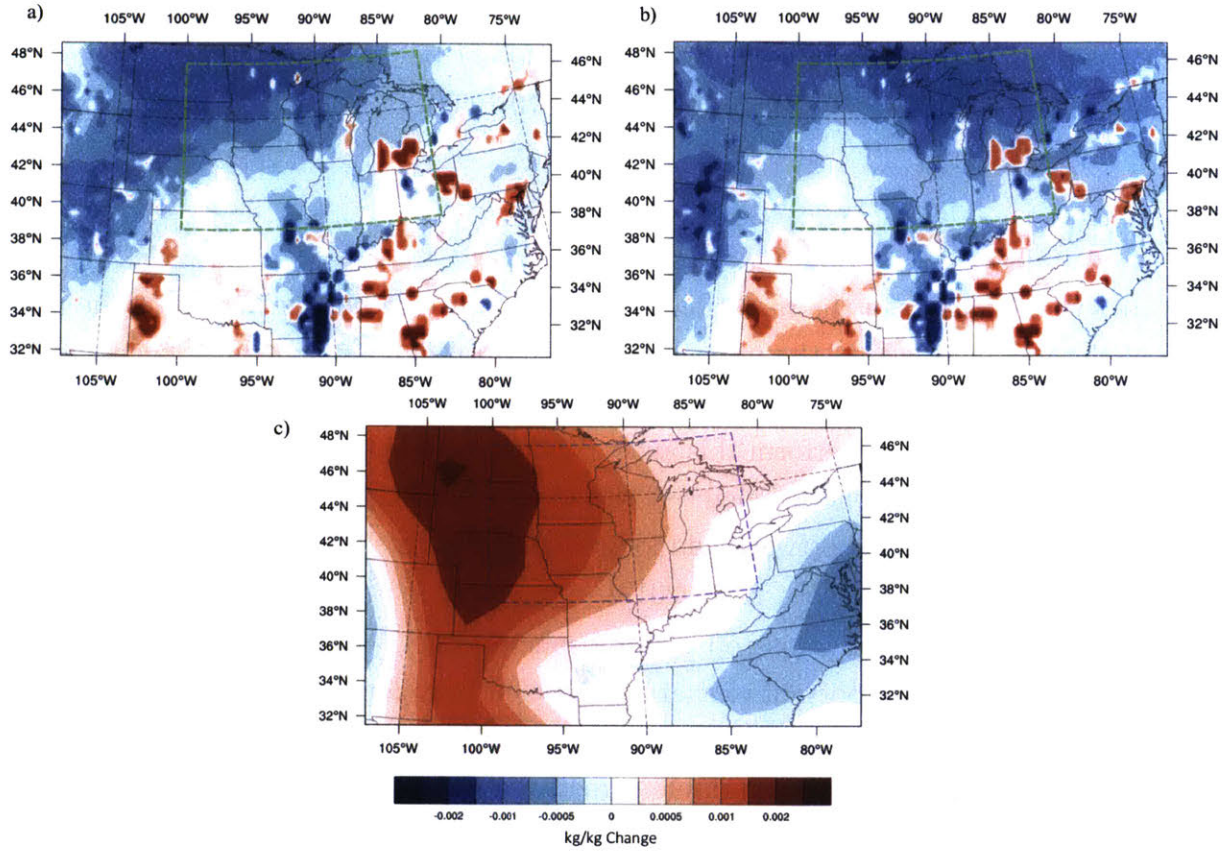


Figure 5-7: July-August surface specific humidity change (kg/kg) from 1920-1949 to 1970-1999 in a) CERA20C VG b) ERA20C VG c) NOAA 20CR V2c. The dotted boxes denote the ROSC.

The early to late period changes in NOAA, ERA20C, and CERA20C in the ROSC are 0.00087, -0.00058, and -0.00043 kg/kg respectively. The magnitude of the change in specific humidity is larger than that shown in the VG simulations and also of opposite sign. The full model simulations show a drying while the observations show an increase in humidity over

this period. For the domain, these corresponding averages are 0.00056, -0.00033, and -0.00023 kg/kg exhibiting the same relationship. Therefore, in both cases the CERA20C simulation specific humidity changes comes closer to the NOAA reconstruction. The changes in specific humidity were significant for all three datasets in the ROSC ($p = 0.0259$, $p = 0.0017$, and $p = 0.0259$) as well as the domain ($p = 0.0017$, $p = 0.0259$, and $p = 0.0259$).

The NOAA data shows a center of increase in specific humidity directly to the west of the ROSC, where the model results show decreases in that same area. All datasets show an increase in specific humidity in the southern portion of the domain (TX in particular) although the model results show a smaller magnitude of change.

The increases in specific humidity in the NOAA data are also matched by specific humidity analysis in Douglas (2016) where ISD station data showed a general increase in specific humidity, although the length of record at the 12 stations used mean that the trends were only analyzed for the mid-century onward and would not be representative of changes from the early period (1920-1949) to the late period (1970-1999).

5.1.4 Radiation Budget

There are several aspects of the radiation budget that can add further insight to the performance of the model. Because the impacts of agricultural development have been shown to stimulate evapotranspiration and downwind rainfall, cloud cover over these regions is likely to be impacted. In order to perform this analysis, data were obtained from the NASA Langley Research Center Atmospheric Sciences Data Center NASA/GEWEX SRB Project which provides monthly values of shortwave and longwave parameters from July 1983 to December 2007 at 1° by 1° resolution (SRB, 2010, 2012).

The first variable looked at is top of atmosphere (TOA) outgoing longwave radiation (OLR), a variable that is important to characterize cloud development and convection. Lower OLR values are associated with higher, and colder, cloud tops (Peixoto & Oort, 1992). Here, because the model provides this variable directly (FIRTP) a direct comparison of OLR between both the model simulations and the NASA data is possible. OLR comparisons are made in the ROSC (Figure 5-8) because this area encompasses the region of greatest agricultural productivity growth over the full 20th century. The average error in the ERA20C simulation is an underestimation of 2.69 W/m^2 while in CERA20C this underestimation is 3.24 W/m^2.

The second measure of interest is the planetary albedo, also a function of cloud cover (Figure 5-9). It a measure of the reflected solar radiation relative to the incoming solar radiation at the top of the atmosphere (Peixoto & Oort, 1992). Because the current setup

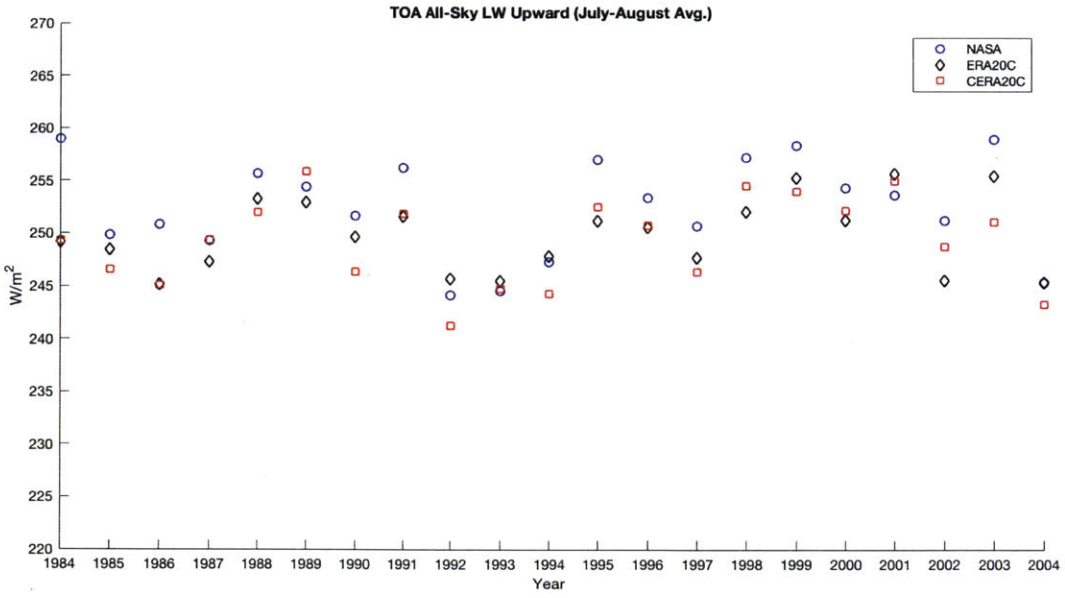


Figure 5-8: July-August average of all-sky outgoing longwave top of atmosphere radiation for NASA-SRB (blue circles), ERA20C (black diamonds) and CERA20C (red squares).

of MRCM does not provide outgoing TOA shortwave radiation, planetary albedo is approximated using $(1 - (OLR/SOLIN))$ where $SOLIN$ is given in the model as the incoming TOA solar radiation and is also provided as a parameter in the NASA SRB dataset (Gianotti, 2013). It is important to maintain the same formulation for planetary albedo across both datasets. This is because, while there is ultimately a balance between incoming SW and outgoing SW + LW at TOA, this occurs over yearly timescales, and the net energy flux can be highly variable at mid latitudes from month to month (Peixoto & Oort, 1992). The average error in the ERA20C simulation is an underestimation of -0.0125 while in CERA20C this underestimation is -0.0111.

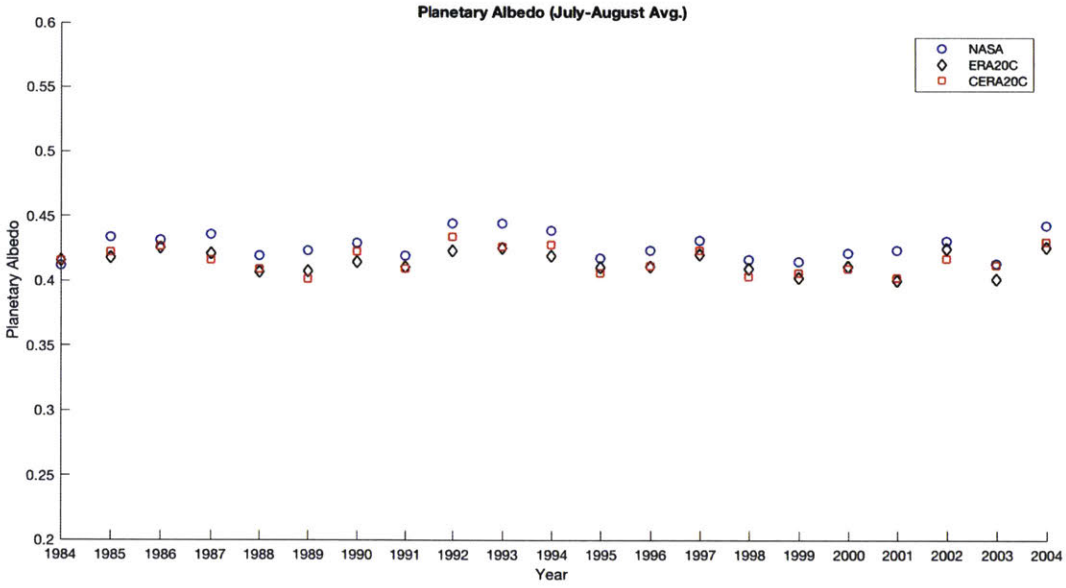


Figure 5-9: July-August average of planetary albedo for NASA-SRB (blue circles), ERA20C (black diamonds) and CERA20C (red squares).

5.2 Comparison to Illinois State Water Survey

While gridded observational datasets allow for comparison at region-wide scales, smaller observational networks are useful for more rigorous study of a small area, and provide valuable information about variables that have high spatial heterogeneity and difficulty of measurement such as soil moisture and evapotranspiration. While the focus of this study is changes in summer climate, climatologies for the full year are shown here.

The Illinois State Water Survey (ISWS), managed by the University of Illinois at Urbana-Champaign, is one of the most comprehensive sets of data that fully characterizes the water budget in a given area (more information and data can be found at https://www.isws.illinois.edu/). In this study, a subset of the data is used that can be compared with the 1901-2005 simulations, and that has also been used and validated extensively in previous studies (Winter, 2010; Winter & Eltahir, 2012a; Winter et al., 2015; Yeh et al., 1998). Total precipitation is averaged from 117 ISWS stations throughout the state. Soil moisture data is measured at 19 Illinois Climate Network (ICN) stations (Figure 5-10) across 11 layers down to 2m (Hollinger & Isard, 1994). Sixteen of the nineteen stations that had data available as far back as 1984 are used, with two stations at Dixon Springs. Runoff was calculated using area weighted streamflow from gages at the Illinois River at Valley City, Rock River near Joslin, and Kaskaskia River near Venedy. Their combined drainage area covers approximately two thirds of Illinois (Yeh et al., 1998). Illinois is also at the center of the simulation domain,

and covered mostly in cropland, providing a test for the vegetation-related changes that were made. Although the ISWS data was the main source of observational data, temperature and precipitation observations were supplemented using the CRUTS4.01 and UDELv4.01. Analyses presented here are made using the CERA20C VG simulation.

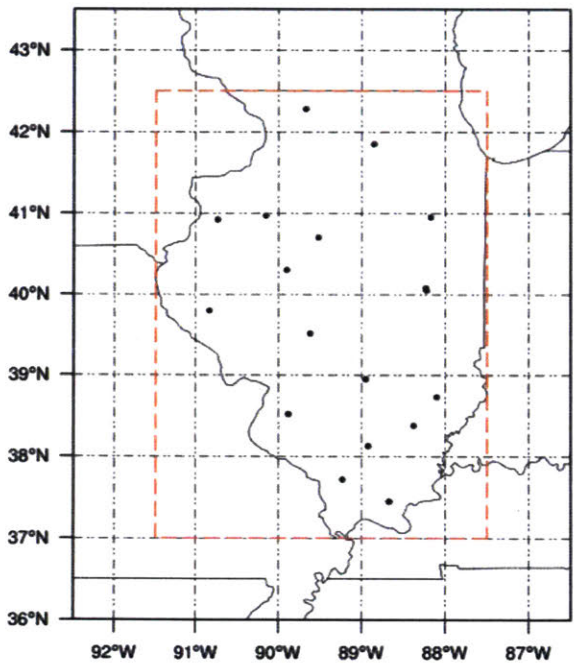


Figure 5-10: Location of ISWS soil moisture measurement stations are marked. The red box shows the area over which model and observational averages were made. The region for averaging gridded data was 37 to 42.5° North and 87.5 to 91.5° West.

Using available parameters from 1983-1994, Yeh et al. (1998) showed that the data from the ISWS could be used with both a surface water balance and an atmospheric water balance approach to estimate the evapotranspiration in the area. This is important, as direct measurements of evapotranspiration cannot be made across large regions, and therefore this study provides a point of comparison akin to observations.

We first begin by quantifying the water budget over Illinois, and making a comparison to the findings in Yeh et al., (1998). The soil water budget can be written as follows (from Yeh et al., 1998):

$$n\bar{D}\frac{\partial \bar{s}}{\partial t} = \bar{P} - \bar{E} - \bar{R_S} - \bar{P_G}$$

Where the term on the left-hand side describes soil water storage, P is precipitation, E is evapotranspiration, R_S is surface runoff and P_G is percolation into groundwater. P_G can

be further decomposed as (from Yeh et al., 1998):

$$\bar{S}_y \frac{\partial \bar{H}}{\partial x} + \bar{R}_G = \bar{P}_G$$

Where P_G is the sum of groundwater discharge (R_G) and storage. Because MRCM does not simulate water table dynamics, we can assume that $P_G = R_G$. Model output of groundwater discharge was not available, so an approximation was made using precipitation, evapotranspiration, surface runoff, and changes in subsurface storage by month along with total yearly amounts, assuming that changes in soil moisture are negligible over the totality of the year. This allowed for calculation of the total runoff which will be used from this point on in the analysis.

A comparison of precipitation, evaporation, and runoff climatology between the model and Yeh et al. (1998) is shown in Figure 5-11. The model overestimates precipitation in the spring and summer and underestimates in the winter. It similarly overestimates evapotranspiration in the winter and spring, and underestimates in the early summer, with good agreement in July through October. Because this is the crop growing season in the model, it adds validation to the adjustment of the photosynthesis parameters to model crop yield. The overestimation of winter and spring evapotranspiration is a feature that has been found in other IBIS/Arakawa-Schubert simulations with MRCM in the past (Winter & Eltahir, 2012a). The model tracks the general shape of runoff throughout the year, but tends to overestimate, especially in the summer months.

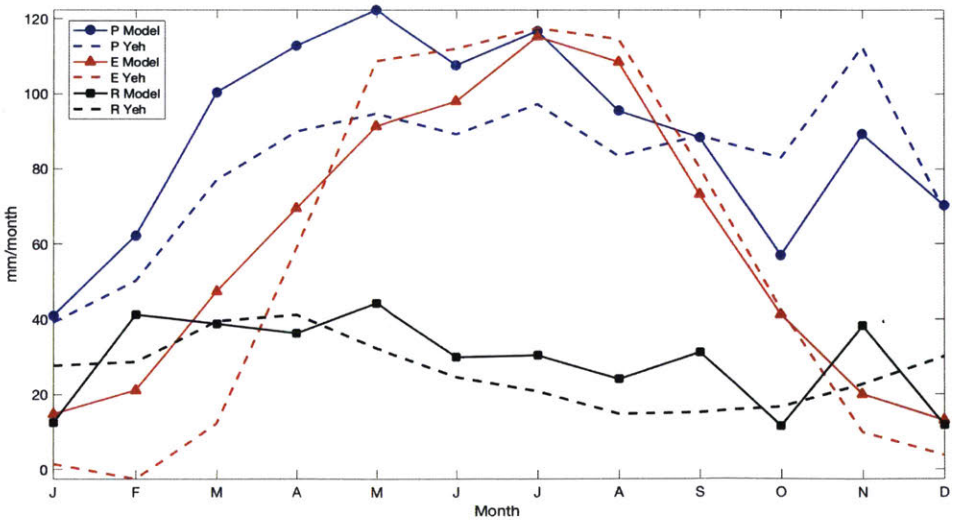


Figure 5-11: Comparison of CERA20C 1983-1994 precipitation and evaporation climatology in CERA20C, and values derived and used in Yeh et al. (1998)

A table quantifying the seasonal and annual differences between the model output and the Yeh et al. (1998) analysis can be seen in Tables 5.1 & 5.2. The four traditional seasons were analyzed as well as July-October for the growing season of corn. Over the 1983-1994 period, the model on average overestimates annual precipitation, evapotranspiration, and runoff by 9%, 8%, and 11% respectively.

Table 5.1: Seasonal water budget values compared between CERA20C VG and Yeh et al. (1998). Positive values indicate higher values in the model

	Avg. P (mm/month)	P %	Avg. E (mm/month)	E %	Avg. R (mm/month)	R %
MAM	24.58	28.16%	9.47	15.8%	2.14	5.69%
JJA	16.59	18.42%	-7.45	-6.5%	8.06	40.13%
SON	-16.6	-17.50%	0.506	1.14%	8.79	48.39%
DJF	5.04	9.54%	14.56	824.40%	-7.00	-24.32%
JASO	1.19	1.35%	-4.24	-4.78%	7.44	44.11%

Table 5.2: Annual water budget values compared between CERA20C VG and Yeh et al. (1998)

	P (mm/yr)	E (mm/yr)	R (mm/yr)
CERA20C VG	1063.8	713.57	350.18
Yeh et al. (1998)	974.8	659.8	314.1

Now wider comparisons are conducted with observational data available in Illinois for temperature, precipitation, surface soil moisture, root zone soil moisture, and runoff. Two-meter surface temperatures from 1970 to 2004 were compared to CRUTS4.01 and UDELv4.01 (Figure 5-12). Model temperatures were warmer than both sets of observations for the first half of the year (1° and 4° C respectively on average), and consistently warmer than the UDEL dataset for the whole year. The model was on average about 2° C cooler than CRU observations in the late summer and fall, but was within 1° C of UDEL observations during this period.

Precipitation from 1970-2004 was compared to CRUTS4.01, UDELv4.01 and the ISWS station network data (Figure 5-13). ISWS data very closely matched CRU data over this period, and therefore errors were only analyzed between CRU and UDEL for this period. The model has a tendency to overestimate precipitation totals in the spring and summer by 20-30 mm/month and underestimate precipitation in the fall by 10-20 mm/month. This is

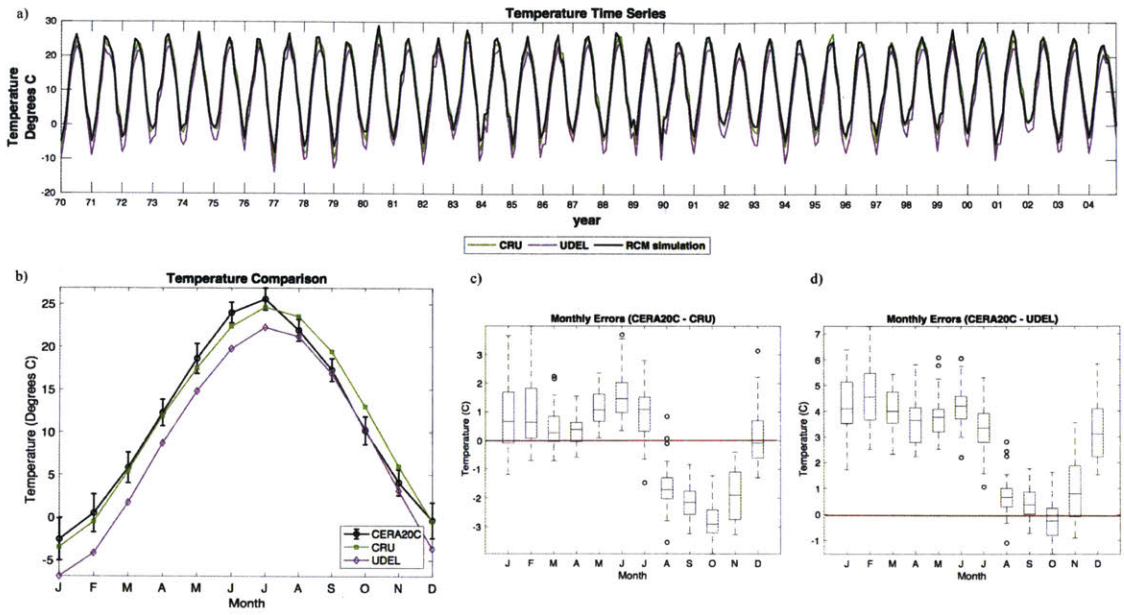


Figure 5-12: 2m Temperature comparisons between model (CERA20C) and observations (CRU, UDEL). a) time-series comparison from 1970-2004 b) yearly climatology of means, with bars on CERA20C showing the standard deviation of monthly values for the period c) quantification of errors between CERA20C and CRU with box showing median, 25% and 75% percentiles, and whiskers showing range of data. Open circles show outliers d) quantification of errors between CERA20C and UDEL with box showing median, 25% and 75% percentiles, and whiskers showing range of data. Open circles show outliers.

a seasonal bias that has been shown with the IBIS/Arakawa-Schubert configuration in the past. The model tracks the general shape of precipitation over the year but exaggerates it by overestimating precipitation in wet months and underestimating it in relatively dry months.

Surface soil moisture, defined as the layer 0 - 10 cm, was compared to data averaged over 16 Illinois Climate Network stations from 1984-2004 (Figure 5-14). The simulations show similar dry biases in the winter and spring and wet biases in the summer and early fall. This wet bias is likely due to overestimation of precipitation in the model in spring and summer and underestimation of evapotranspiration. An overestimation of evapotranspiration in the winter and spring likely leads to the dry bias during those periods. As can be seen in the full time-series, the model generally misses the peaks and troughs of surface soil moisture in a given year. For example, the model misses the low soil moisture observed in the drought of 1988 and the high saturation level from the floods of 1993 despite the fact that precipitation matched the observations well during these periods.

Root zone soil moisture, defined as the layer 0 - 100 cm, was also compared to data averaged over 16 Illinois Climate Network stations from 1984-2004 (Figure 5-15). ICN data from the first 6 layers (0 - 110 cm) was weighted by layer thickness to represent total soil

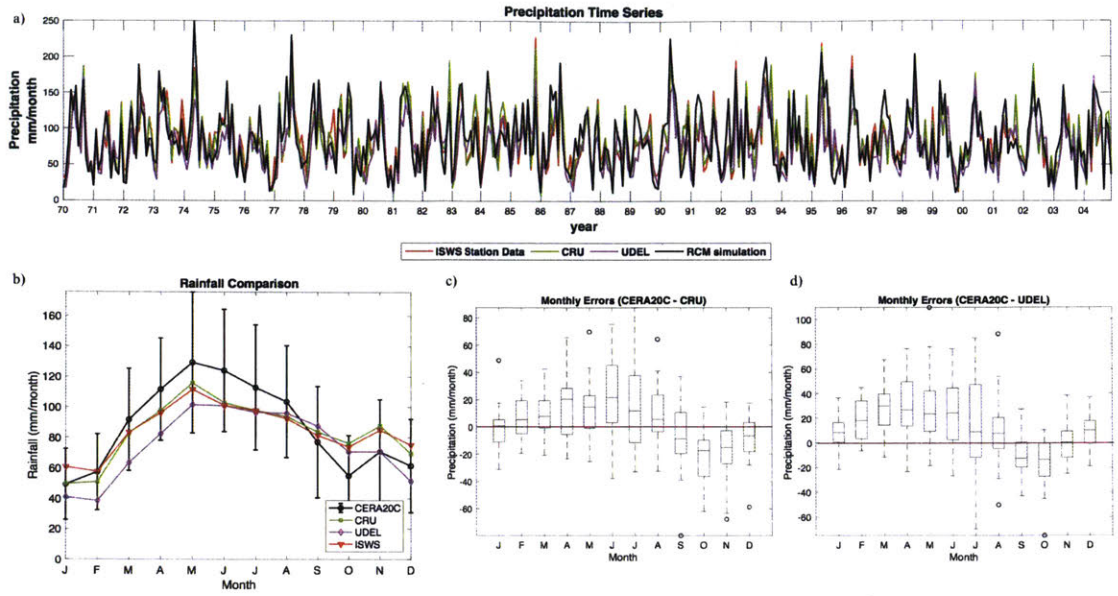


Figure 5-13: Precipitation comparisons between model (CERA20C) and observations (CRU, UDEL, ISWS). a) time-series comparison from 1970-2004 b) yearly climatology of means, with bars on CERA20C showing the standard deviation of monthly values for the period c) quantification of errors between CERA20C and CRU with box showing median, 25% and 75% percentiles, and whiskers showing range of data. Open circles show outliers d) quantification of errors between CERA20C and UDEL with box showing median, 25% and 75% percentiles, and whiskers showing range of data. Open circles show outliers

moisture within the root zone, and therefore it is possible that soil saturation would be slightly higher in this thicker layer than it would be in the true 100 cm. Root zone soil moisture shows a dry bias throughout much of the year but matches observations well in the summer months. This bias towards root zone dryness has been observed in other studies using the IBIS/Arakawa Schubert setup, although the current model improves on the deficit in these summer months (Winter & Eltahir, 2012a). Winter and Eltahir (2012a) also found that the minimum for root zone soil moisture was delayed one month in summer from observations, an important feature present in these results. Not only is this minimum delayed, it is on average lower than the one observed in the observational data.

Runoff was compared to drainage area weighted stream flow from 1984-2004 for three ISWS network gages that capture two-thirds of the drainage area of Illinois (Figure 5-16). There was less seasonal consistency in the error of runoff, but in general runoff was overestimated in the summer and fall. This overestimation may be in part because of the larger assumed drainage area in the model versus the observations. It is unclear whether the model was overestimating surface runoff or baseflow since the observational data did not partition between the two. Yeh et al. (1998) found that while baseflow was the main mode of runoff at

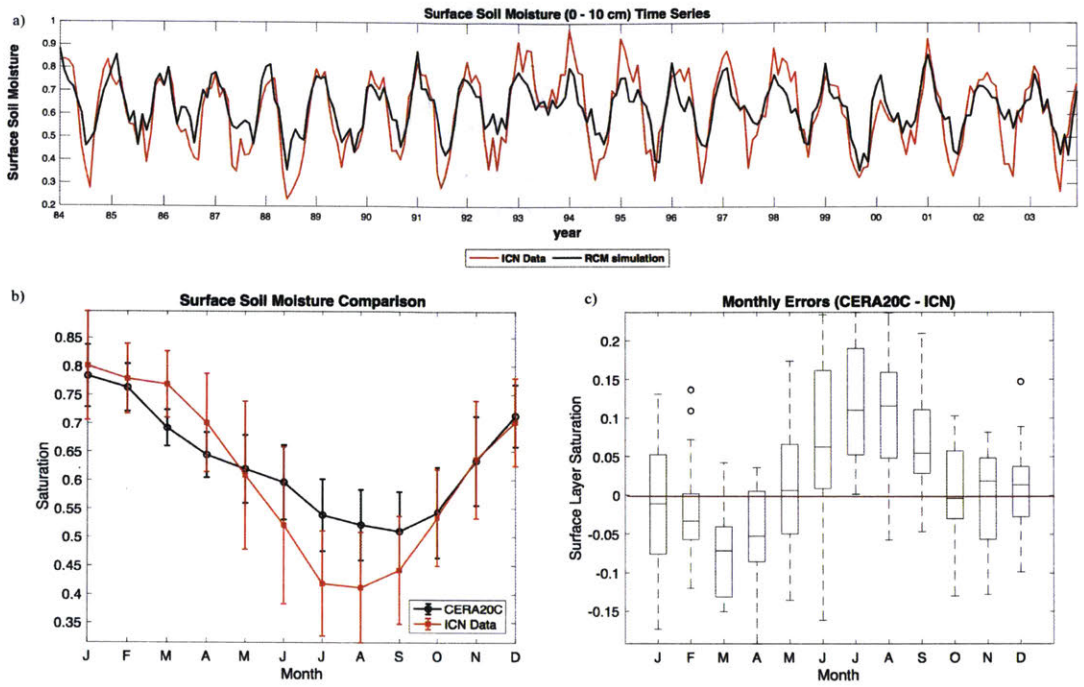


Figure 5-14: Surface soil moisture saturation comparisons between model (CERA20C) and observations (ICN) a) time-series comparison from 1984-2004 b) yearly climatology of means, with bars showing the standard deviation of monthly values for the period c) quantification of errors between CERA20C and ICN stations with box showing median, 25% and 75% percentiles, and whiskers showing range of data. Open circles show outliers.

this site, many models had difficultly correctly partitioning runoff, and generated significant levels of overland (surface) runoff.

This simulation exhibits some of the same tendencies present in other Midwest hydroclimatological studies with the RegCM3-IBIS Arakawa Schubert setup, but is improved in some respects, particularly soil moisture. Overall, MRCM overestimates spring and summer rainfall by roughly 25%, underestimates summer evapotranspiration by about 7%, and overestimates runoff by as much as 50%. Consequently, MRCM overestimates summer surface soil moisture, but tends to underestimate summer root zone soil moisture to a lesser degree, and is an improvement on past simulation results in this area.

Importantly for plant water availability in the growing season, the summer minimum in root zone soil moisture (1984-2004 avg.) is overestimated and delayed by a month, from August to September. However, the model average shows fewer months in the climatology in which evapotranspiration exceeds precipitation on average versus estimations in Yeh et al. (1998). This is likely due to deficiencies in the soil moisture parameterization with respect to groundwater interactions.

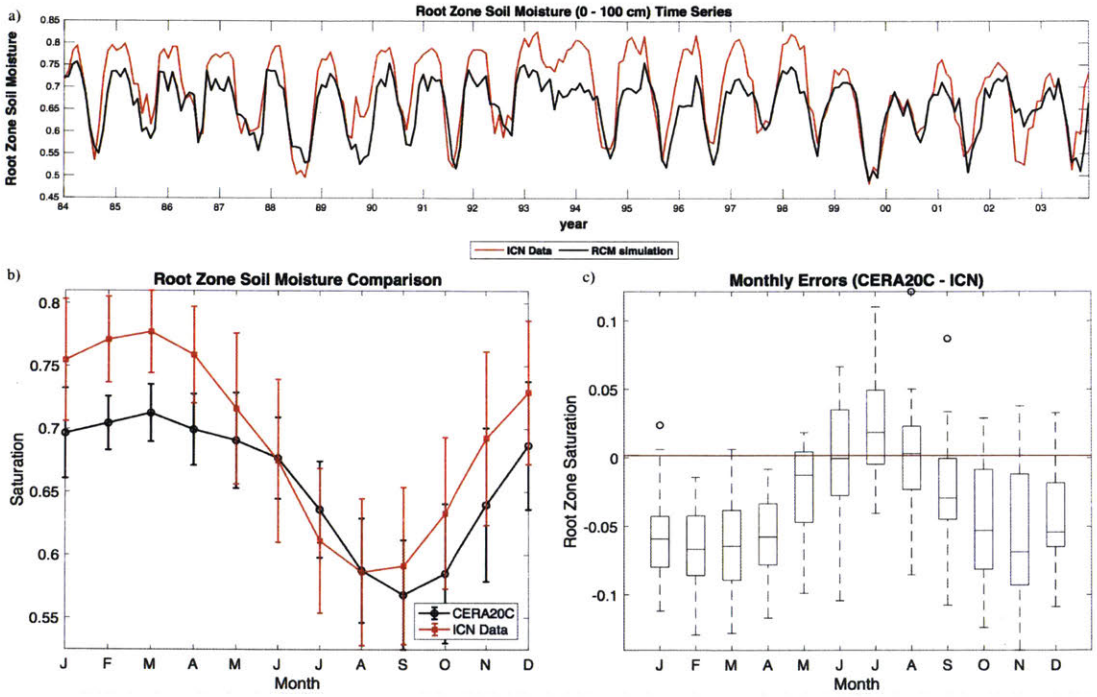


Figure 5-15: Root Zone soil moisture saturation comparisons between model (CERA20C) and observations (ICN) a) time-series comparison from 1984-2004 b) yearly climatology of means, with bars showing the standard deviation of monthly values for the period c) quantification of errors between CERA20C and ICN stations with box showing median, 25% and 75% percentiles, and whiskers showing range of data. Open circles show outliers

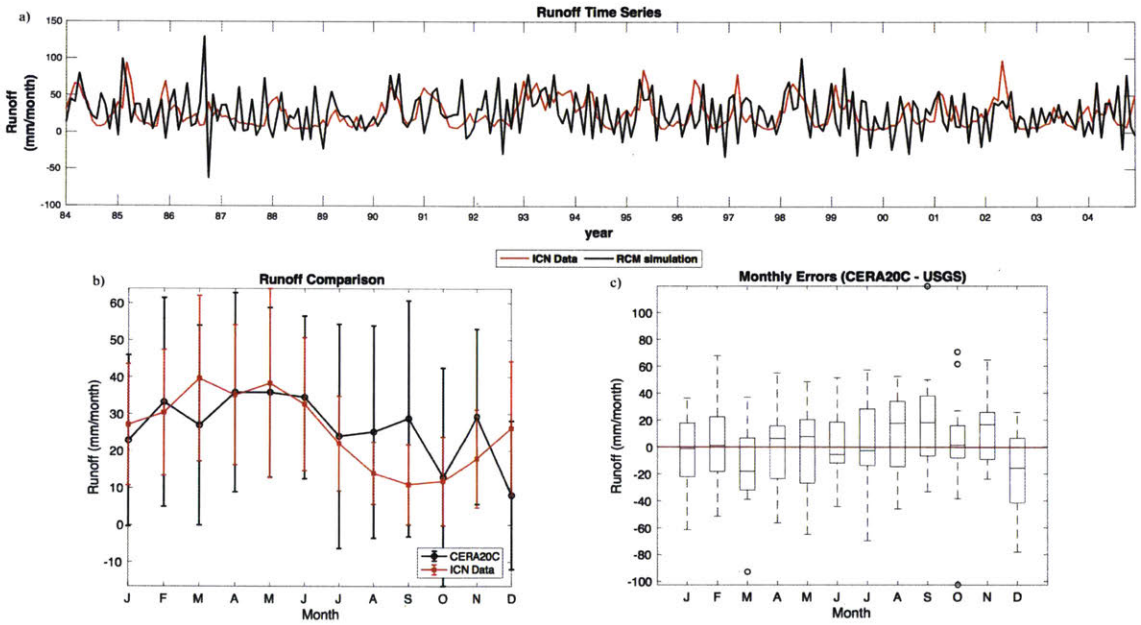


Figure 5-16: Runoff comparisons between model (CERA20C) and observations (ISWS Gages) a) time-series comparison from 1984-2005 b) yearly climatology of means, with bars showing the standard deviation of monthly values for the period c) quantification of errors between CERA20C and ISWS streamflow with box showing median, 25% and 75% percentiles, and whiskers showing range of data. Open circles show outliers

5.3 Impacts of Greenhouse Gases

First, this study looks at the effects that greenhouse gas (GHG) evolution within the model has on the climate in this region. GHG are a diffuse forcing and therefore have a diffuse impact with few areas of strong change. A mix of changes are seen within the variables with regional averages that point to the expected effects of GHG increases.

5.3.1 Methods

In order to decompose these impacts, two simulations from each set were used: NVNG and NVG. In both of these, there is no development of vegetation, and the only change is the inclusion of GHG. To show the relative impacts, the variable trends from early to late period in the NVNG simulation were subtracted from the early to late variable trends in the NVG simulations. Therefore, the shown effects are the changes in variables between the early and late period that can be attributed solely to the inclusion of GHG increases in the model.

5.3.2 Analysis

The temperature effect is a general warming interspersed with several pockets of cooling that are more widespread in the south and southeast (Figure 5-17). The average effects for CERA20C and ERA20C are 0.044 and 0.052 °C for the ROSC and 0.025 and 0.036 °C for the domain. Although the overall temperature changes in these regions were negative for the full simulations, the GHG effects caused warming, showing that they were masked by some other forcing mechanism.

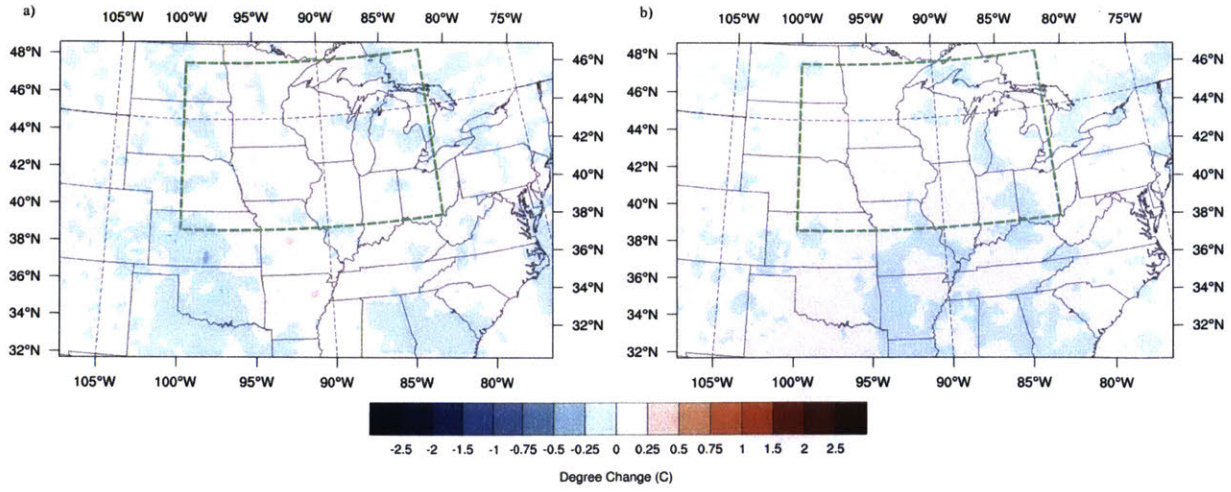


Figure 5-17: July-August average temperature change (°C) from 1920-1949 to 1970-1999 attributed to GHG in a) CERA20C b) ERA20C

The precipitation effects are similarly scattered in the domain (Figure 5-18). The average GHG related precipitation changes for CERA20C and ERA20C are 0.021 and 0.017 mm/day for the ROSC and 0.026 and 0.032 mm/day for the domain. Alter et al. (2015a) notes that greenhouse gas induced climate change will tend to cause more frequent, heavy precipitation events while maintaining the overall precipitation totals, and therefore the results here are consistent.

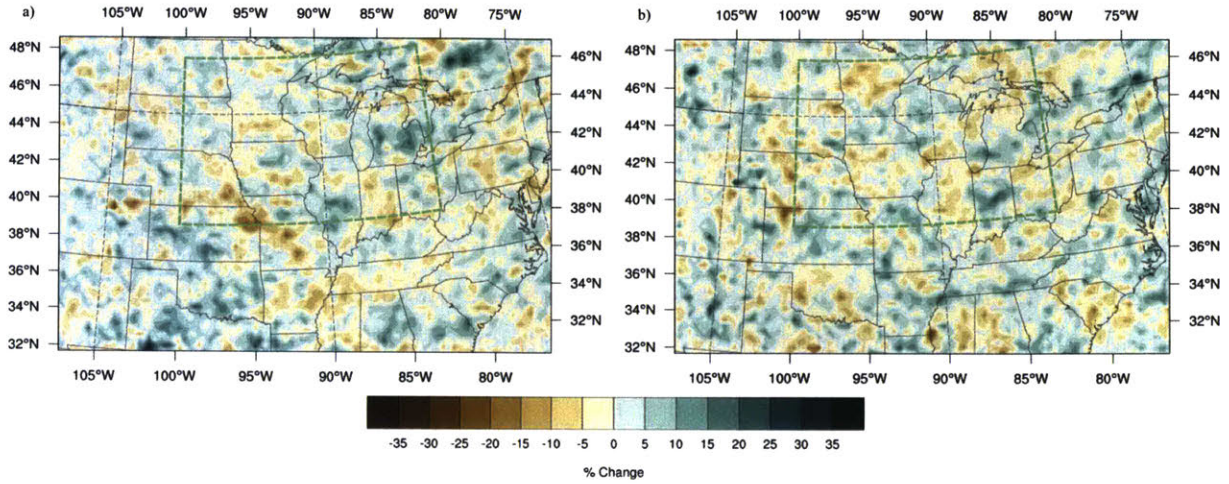


Figure 5-18: July-August average daily precipitation change (%) from 1920-1949 to 1970-1999 attributed to GHG in a) CERA20C b) ERA20C

A similar increase exists for specific humidity (Figure 5-19). The average GHG related changes for CERA20C and ERA20C are 0.000024 and 0.000018 kg/kg in the ROSC and 0.000028 and 0.000027 kg/kg in the domain.

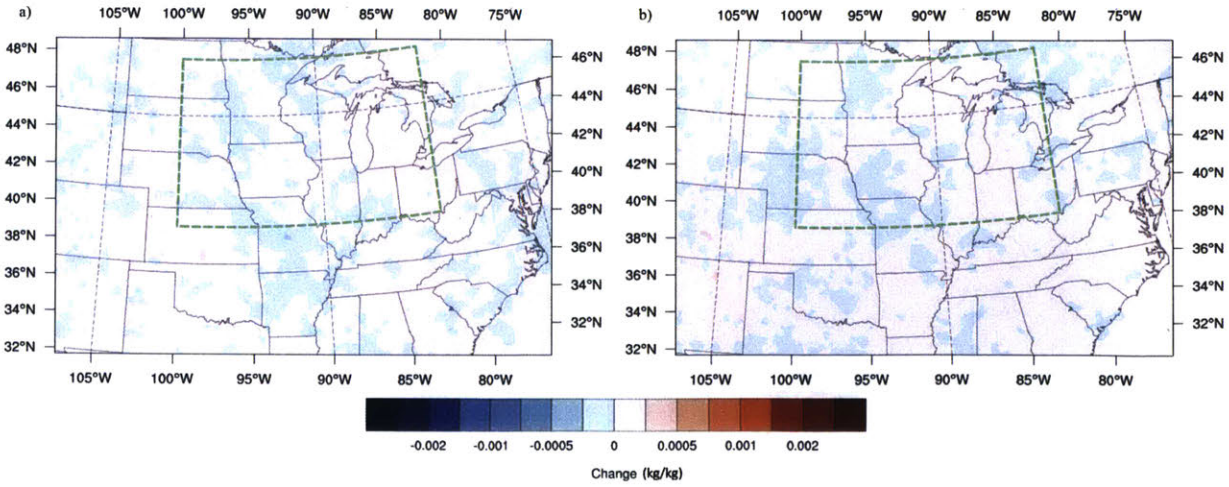


Figure 5-19: July-August average specific humidity change (kg/kg) from 1920-1949 to 1970-1999 attributed to GHG in a) CERA20C b) ERA20C

The model seems to have a more heterogeneous effect on evapotranspiration due to GHG forcing than in most of the other variables (Figure 5-20). This may be due to the presentation of the data in percentage change, where similar increases in evapotranspiration are shown to be a larger fraction of the base amount. Overall, the absolute GHG effect on evapotranspiration is still small, with average changes in CERA20C and ERA20C of 0.0080 and 0.0010 mm/day in the ROSC and 0.017 and 0.0095 mm/day in the whole domain.

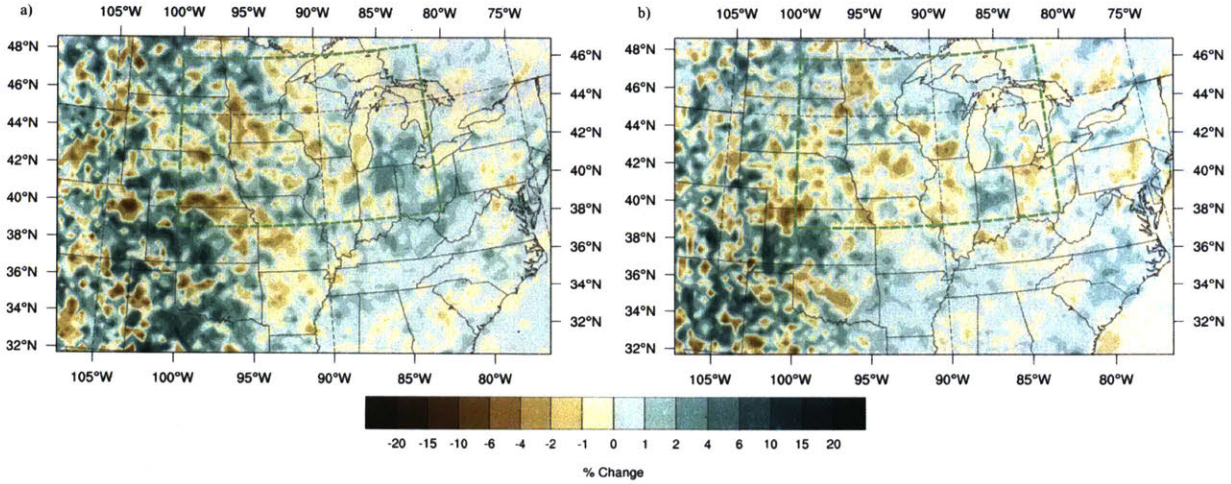


Figure 5-20: July-August average daily evapotranspiration change (%) from 1920-1949 to 1970-1999 attributed to GHG in a) CERA20C b) ERA20C

Finally, the effects of GHG on the other element of the surface energy balance, mainly sensible heat flux, are examined (Figure 5-21). As expected, the areas that show cooling due to greenhouse gases within either of the simulations also see a decrease in sensible heat flux. The average changes in CERA20C and ERA20C are 0.20 and 0.27 W/m^2 in the ROSC and 0.024 and 0.15 W/m^2 in the domain. On the whole, the average change in sensible heat is positive due to the overall warming effect.

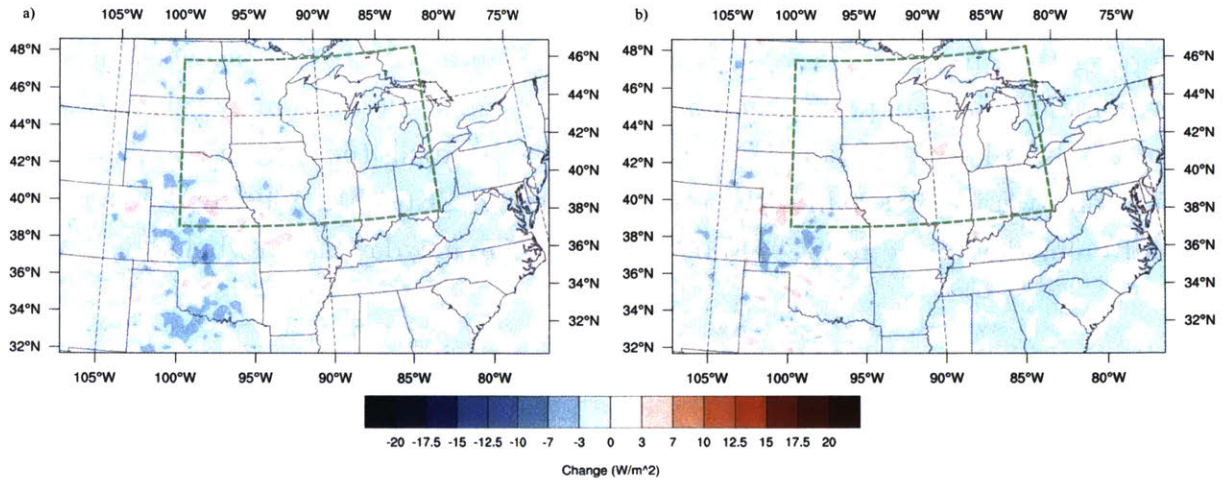


Figure 5-21: July-August average sensible heat change (W/m^2) from 1920-1949 to 1970-1999 attributed to GHG in a) CERA20C b) ERA20C

5.3.3 Discussion

The addition of GHG to the model causes expected changes in the variables, with some local heterogeneity. In all of these variables, GHG contributes only a portion of the observed changes over these regions in the time periods investigated, and that the direction of impact is occasionally counter to that seen in the observations or in the model simulation including all forcings. This is evidence that some climate forcing other than GHG changes was significant in inducing the observed change.

5.4 Impacts of Agricultural Development

Analyzing the impacts of vegetation development in the model is more complicated than GHG for two distinct reasons. First, there are three separate components of agricultural development that were considered and included in the evolution of the model in some way. These are the expansion of cropland, increases in agricultural productivity, and the expansion of irrigated areas. All of these components have different impacts and may work simultaneously in the domain, especially with regard to expansion of cropland types and corresponding intensification.

Secondly, the development of agriculture is heterogeneous in space and interacts with the natural vegetation in a way that is unique to a particular grid cell. This means that the development of cropland from natural vegetation or reversion back into natural vegetation is individual to each grid cell. As noted in Chapter 1, the potential vegetation that existed before the development of cropland is important for the exact nature of the climate impact,

and these sets of pre- and post-development vegetation pairs vary throughout the domain. Additionally, in some areas vegetation type changed dramatically and completely from the early to late period, showing sharp areas of intense change that can be seen mostly in the east and in the irrigated areas. The direction of these changes is dependent on the grid cells natural vegetation counterpart.

5.4.1 Methods

A similar methodology is used here to isolate the effects of vegetation developments and their impact on this study's five variables of choice. Because all three agricultural developments were tested in an amalgamation in the model, their effects cannot be separated on a region-wide scale. However, the local effects can be examined where certain changes have been known to have taken place.

In order to decompose these impacts two simulations from each set were used: NVG and VG. In both of these, there are changes in GHG, and the only difference between the two simulations is the inclusion of vegetation in the form of updated land use datasets to the surface boundary and increasing multiplication factors of photosynthesis rate (discussed in Chapter 4). To show the relative impacts, the variable trends from early to late period in the NVG simulation were subtracted from the early to late variable trends in the VG simulations. Therefore, the shown effects are the changes in variables between the early and late period that can be attributed solely to the inclusion of developing vegetation in the model.

5.4.2 Analysis

The average temperature changes in CERA20C and ERA20C are -0.32 and -0.3 °C in the ROSC and -0.18 and -0.17 °C in the domain (Figure 5-22). Most of this cooling is concentrated in areas of agricultural development in the Midwest, and punctuated by strong cooling in late century irrigated areas. While the areas that were maintained as cropland or that saw steady expansion have smoother cooling features and less intense cooling, some areas show up as vibrant splotches as mentioned earlier. The warming areas occur mostly due to the re-emergence of natural vegetation in north Texas and Oklahoma, and in the east.

The vegetation effects are also lacking the strong anomalous cooling in the northwest and the strong warming in the southeast that was seen in the analysis of the full simulation results in an earlier section (Figure 5-3).

The average precipitation changes in CERA20C and ERA20C are 0.088 and 0.067 mm/day in the ROSC and 0.043 and 0.043 mm/day in the domain (Figure 5-23). Much like in the

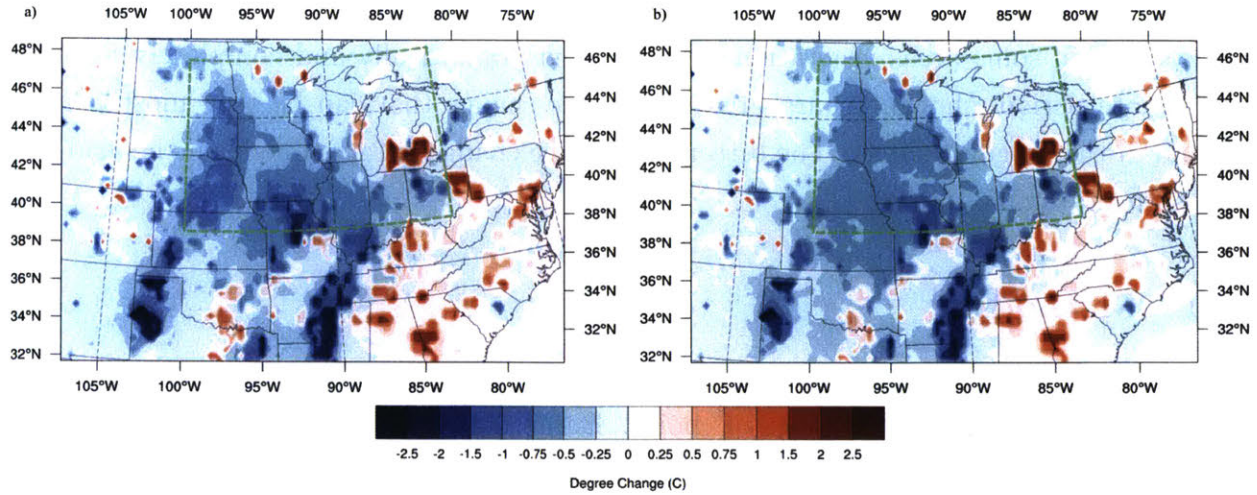


Figure 5-22: July-August average temperature change (°C) from 1920-1949 to 1970-1999 attributed to Vegetation Development in a) CERA20C b) ERA20C

case of GHG effects, the map of vegetation impacts shows almost no distinct features of precipitation increases or decreases. In CERA20C there seems to be a region of increased precipitation at the southern edge of the ROSC and in the ERA20C there are a few features of increased precipitation in the north and east of the ROSC that coincide with precipitation increases in the observations. There does not seem to be any relationship between locations of agriculture development and increased precipitation, nor with irrigation specifically. Precipitation is one of the variables that would most greatly benefit from a composite result made of multiple simulation ensembles. The results presented here for ERA20C and CERA20C are the results of a single run for each decade that makes up the composite century.

The average specific humidity changes in CERA20C and ERA20C are 0.00027 and 0.00023 kg/kg in the ROSC and 0.00019 and 0.00016 kg/kg in the domain (Figure 5-24). Intense agriculture and irrigation here have the expected effect of increasing low-level atmospheric moisture content. This is a very different map of the early to late period change than was seen in the full simulation comparison to observations. In this case, there is an increase in specific humidity across most of the agricultural areas, save for one region of comparatively strong drying in the Mississippi River irrigation area.

The average evapotranspiration changes in CERA20C and ERA20C are 0.16 and 0.15 mm/day in the ROSC and 0.10 and 0.09 mm/day in the domain (Figure 5-25). Again, the areas that exhibit the most substantial changes are those that are covered in cropland, and particularly irrigated areas, where the change in daily evapotranspiration can exceed 20% and is double the increase that occurs for non-irrigated cropland. Note that the scale on the evapotranspiration is not uniform in order to show gradations in the areas with large in-

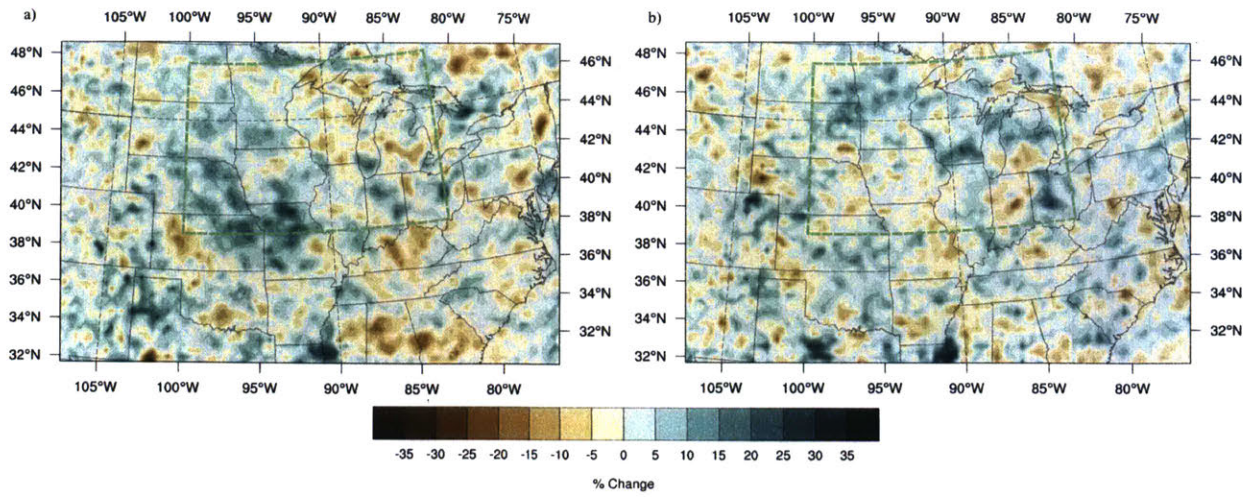


Figure 5-23: July-August average daily precipitation change (%) from 1920-1949 to 1970-1999 attributed to Vegetation Development in a) CERA20C b) ERA20C

creases. The eastern spots are still present here but appear to exhibit a more cohesive trend. These increases in evapotranspiration occur in the east as cropland reverts to forested land type in the model. This increase is consistent with the effect predicted in Twine et al. (2004) and Sterling et al. (2012). In the northwest Midwest, the conversion of grassland type vegetation to cropland would normally imply a decrease of evapotranspiration (Sterling et al., 2012). However, here there is an increase, likely as the increase in agricultural productivity in these croplands makes up for the difference and is consistent with the increases in evapotranspiration expected according to Twine et al. (2004).

In order to show that this increase in ET is linked to an increase in crop productivity — as represented in the model output by net primary productivity — the trends of the two are plotted on the same graph (Figure 5-26). The comparison is made using the ERA20C VG simulation output to compare to the NPP model output which is produced from the 3-hr results given by the model and was taken from the simulation set also using the ERA20C reanalysis. Averages are made over the Prince et al. (2001) region outlined in Chapter 4. Overall, evapotranspiration does not increase to the same degree that NPP does, but beginning around the same time as the NPP increase in the mid 1930s there is a similar linear increase in ET with an interannual variability to match productivity. The anomalously low value of evapotranspiration in 1903 is outside the period of analysis and therefore should not influence results.

Finally, sensible heat flux is expected to decrease in cropland areas to complement the increases in evapotranspiration and surface temperature that have already been seen. As seen in Figure 5-27, The average sensible heat flux changes in CERA20C and ERA20C are -3.72

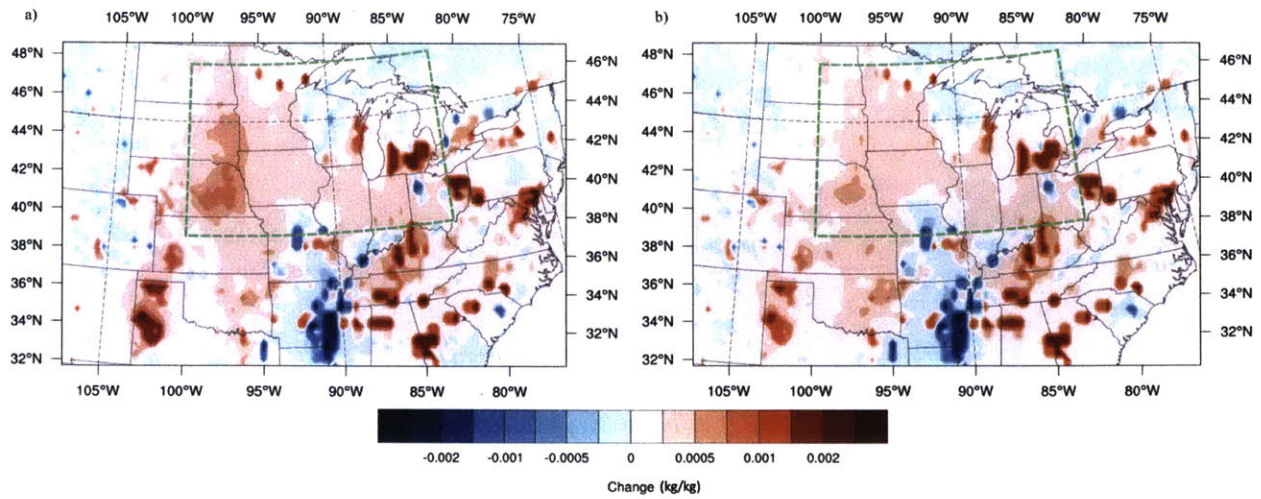


Figure 5-24: July-August average specific humidity change (kg/kg) from 1920-1949 to 1970-1999 attributed to Vegetation Development in a) CERA20C b) ERA20C

and -3.76 W/m^2 in the ROSC and -1.87 and -1.90 W/m^2 in the domain. In comparison to Figure 5-22, the regions of decrease match well, but unlike what is observed in temperature, this reduction does not extend outside of the irrigated areas. The areas in the east where croplands convert from cropland to the natural coniferous forested area exhibit increases in sensible heat, which is the complementary change to the significant decreases in sensible heat that Twine et al. (2004) show occuring in the conversion from conifer and mixed forest to cropland.

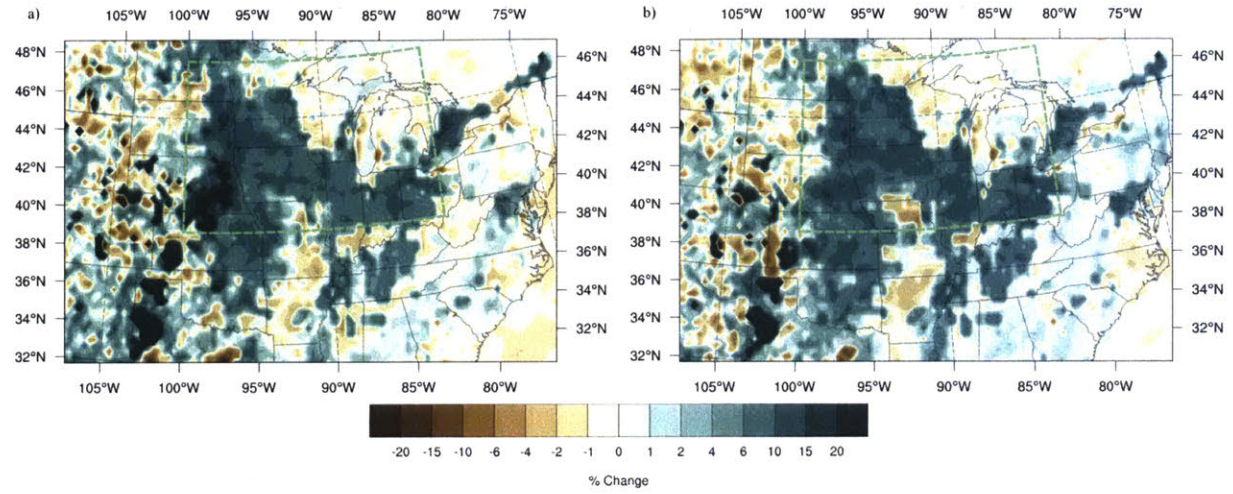


Figure 5-25: July-August average daily evapotranspiration change (%) from 1920-1949 to 1970-1999 attributed to Vegetation Development in a) CERA20C b) ERA20C

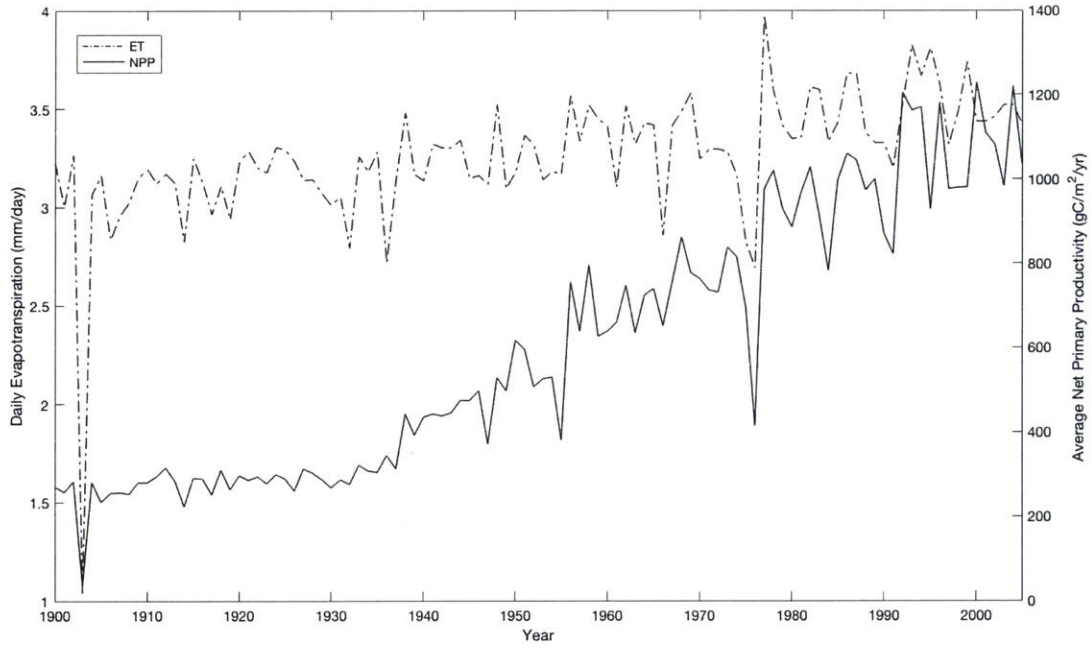


Figure 5-26: Plot of July-August average daily evapotranspiration trend (dotted line) in cropland grid cells, versus the average annual net primary productivity of the same grid cells (solid line).

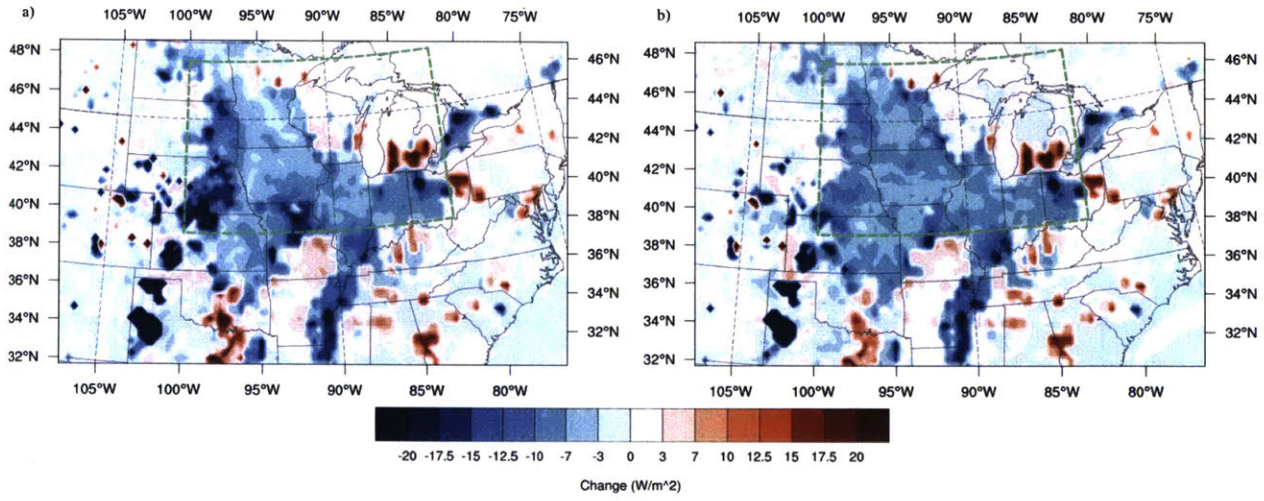


Figure 5-27: July-August average sensible heat change (W/m^2) from 1920-1949 to 1970-1999 attributed to Vegetation Development in a) CERA20C b) ERA20C

It is also illustrative to look at some of the smaller vegetative areas closely to see exactly how varying vegetation dynamics can affect the climate impacts that occur. To that end, a small subset of the full domain is looked at, from 32 – 36°N (Figure 5-28). Within this stretch there are three main features: irrigated areas in the Texas Panhandle (100 – 103°W), irrigation along the Mississippi River (89 – 92°W), and a spot of cropland on the Alabama-Georgia border (85°W).

Irrigation in the Texas Panhandle develops in spots in the 1940s but not substantially until the 1950s and 1960s. Irrigation here replaces grassland in the model, and importantly there is no surrounding non-irrigated cropland. According to Sterling et al. (2012), this transition from grassland to irrigated cropland would cause an increase in evapotranspiration rates even without the accompanying agriculture intensification. This tends to cause sharp variable changes within the irrigated areas that are starkly different from the surroundings. These irrigated areas exhibit the classic behavior that would be expected in irrigated regions. Temperature and sensible heat decrease dramatically, while evapotranspiration increases, along with both surface and root zone soil moisture. The increase in soil moisture is due to the artificial wetting imposed by the irrigation. Specific humidity and precipitation also increase in this region to a lesser degree.

The cropland spot along the Alabama-Georgia border follows an opposite trajectory. Cropland exists in this region at the beginning of the 20th century, and disappears in mid-century. However, instead of the grassland that the TX panhandle irrigation replaced, these areas revert to their designation on the potential vegetation map — temperate evergreen coniferous forest/woodland. These areas show an increase in temperature and associated increase in sensible heat. A decrease in rainfall and an increase in evapotranspiration lead to an increase in specific humidity but a drying of both soil layers. These changes are consistent with those described in Twine et al. (2004).

The irrigated area along the Mississippi River benefits the most from a grid cell by grid cell analysis due to the contrast in changes between neighboring grid cells and the timing of those changes. In this region, there is a mixture of forestland as the potential vegetation type. There is some early replacement of this forest with non-irrigated cropland in the early portion of the century but real development does not come until the late period. Irrigation begins to develop in the 1970s to 1980s along with further expansion of cropland. Extensive irrigation establishes here in the 1990s to 2000s and ultimately this region is a mixture of non-irrigated and irrigated cropland. Because of these decade to decade differences and the averaging that is performed between them, this area has a unique climatological development in the model. Firstly, there is a slight overall decrease in temperature and sensible heat. The impacts on evapotranspiration are mixed, likely due to the forest/cropland mixture and the interplay

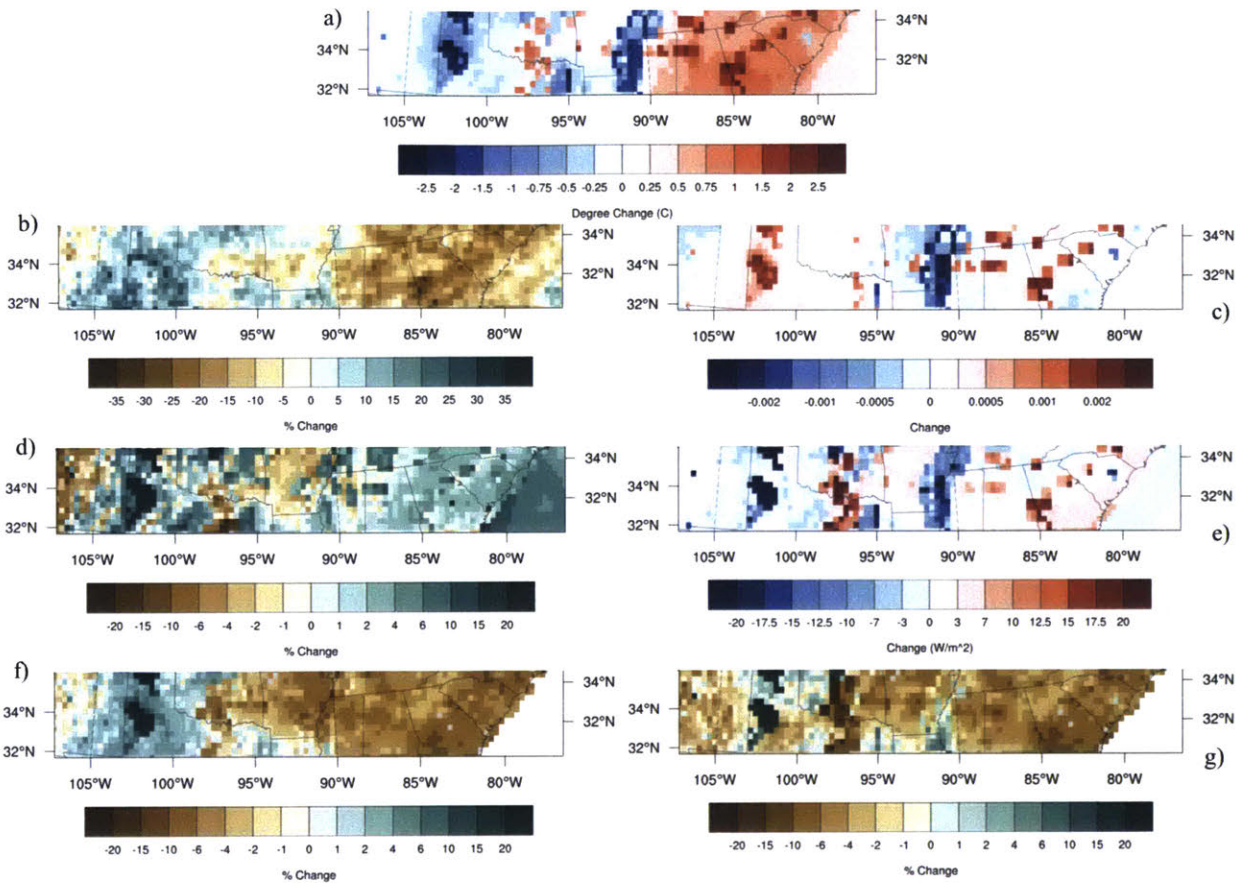


Figure 5-28: July-August average change from 1920-1949 to 1970-1999 in CERA20C VG for selected variables. Results are presented in raster form in order to elucidate the changes occurring at individual pixels for a) temperature (°C) b) precipitation (%) c) specific humidity (kg/kg) d) evapotranspiration (%) e) sensible heat (W/m^2) f) surface soil moisture (0 - 10 cm) (%) g) root zone soil moisture (0 - 100 cm) (%)

of irrigation and rain-fed areas. Similarly, precipitation impacts are mixed. Ultimately this leads to a decrease in specific humidity and mixed impacts in surface soil moisture and root zone soil moisture.

5.4.3 Discussion

In comparison to the effects examined in the GHG section, vegetation development had a stronger impact on the early to late century change of almost all of the considered variables and usually caused an opposite change. However, it is important to consider whether the vegetation changes that were implemented are realistic.

In some ways, the introduction of irrigation into the model is conservative – it does not account for the greater productivity and yield stability that is observed in these areas versus the rain-fed areas. However, the assumption of 25% irrigation density translating to a fully

irrigated grid cell likely overestimates the amount of irrigated land within these pockets in TX, KS, and NE. It should be questioned how much these effects compensate for one another. Even with a supposed higher intensity of irrigation, the expected downwind rain effects were not observed.

The bright spots of change in the Eastern portion of the domain, where early century cropland reverts back to the natural vegetation type in the model, also points to an assumption that may need to be adjusted in the future. There is the question of whether this return to "natural" vegetation would have occurred in the case of forested areas, and how much of this redevelopment would have taken place over the ten-year time span that separates subsequent land surface maps. Ramankutty and Foley (1999) note that abandoned cropland has a tendency not to revert to its full, pre-development vegetation state due to soil degradation and urbanization.

5.5 Background Effects

Finally, this section will examine the background effects due to natural decadal and century scale variability, i.e. the model simulation results un-accounted for by impacts from anthropogenic forcings. This captures natural variability and forcings imposed by reanalysis data. There are many features present in the full model simulations that were not present in either the GHG or vegetation effects and that can be explained by looking at the background effects.

5.5.1 Methods

For this analysis, the NVNG simulation in each set is used, as it contains no greenhouse gas or vegetation development. Both of these forcings are maintained at the 1900 level for the entire simulation. While this does impart a higher than natural level of GHG and a substantial amount of cropland into the model, the goal of this study is to examine the impact of changes over the 20^{th} century and therefore these levels are appropriate. In order to look at the background effects, differences can simply be made between the 1920-1949 and 1970-1999 time periods.

5.5.2 Analysis

The same five variables are described here, along with the notable effects and their influence on the full VG simulations. Firstly, temperature is examined (Figure 5-29). In the NVNG simulations for CERA20C and ERA20C there was an average change of -0.60 and -0.93 °C in

the ROSC and -0.22 and -0.52 °C in the domain as a whole. These changes were significant in the ERA20C simulation in the ROSC (p = 0.0002) and the domain (p = 0.00006). In the CERA20C simulation this change was considered significant in the domain (p = 0.0113) but not in the ROSC (p = 0.2003). The CERA20C background change shows less cooling in the northern portion of the domain as well as in the south, although this northern cooling is shifted to the west. This feature in both simulations is behind the strong cooling in both of the simulations in these regions. Additionally, the heating evident in the southeast in CERA20C and in the west in ERA20C is also evident in the full simulations (VG).

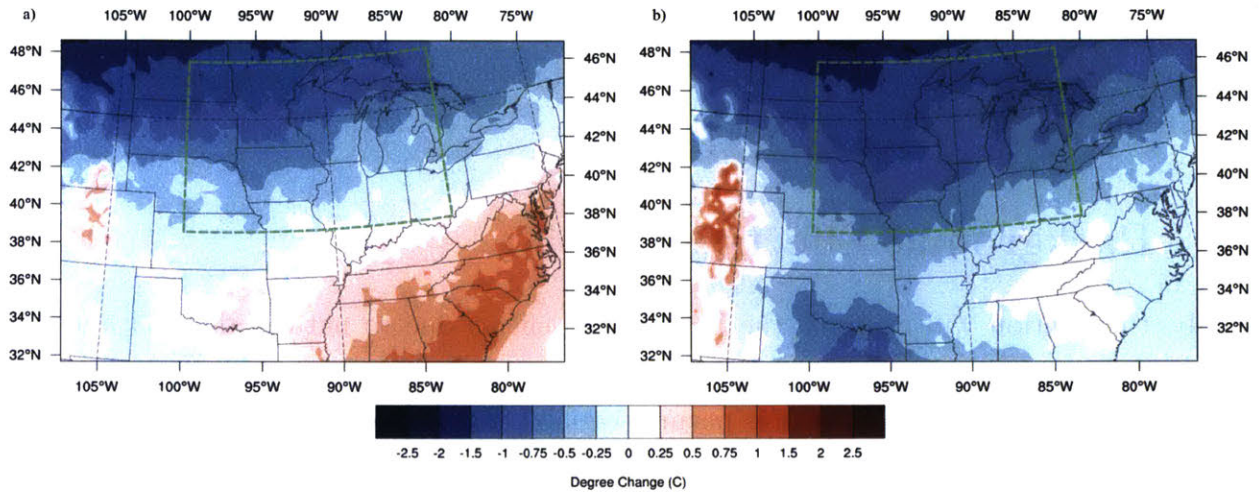


Figure 5-29: July-August average temperature change (°C) from 1920-1949 to 1970-1999 in a) CERA20C NVNG b) ERA20C NVNG

The precipitation patterns seen in the full simulations are also driven largely by background effects (Figure 5-30). This is unsurprising given the relative homogeneity of precipitation changes attributed to either GHG or vegetation. The strong patches of drying are seen here, as well as the band of increased precipitation in the center of the domain. In the NVNG simulations for CERA20C and ERA20C there was an average change of -0.02 and -0.024 mm/day in the ROSC and -0.16 and -0.07 mm/day in the domain as a whole. These changes were not significant in the CERA20C simulation in the ROSC (p = 0.7600) and the domain (p = 0.0550), nor in the ERA20C simulation in the same regions (p = 0.5372 and p = 0.2003).

Again, the anomalous reduction in surface specific humidity is present in the background changes, with drying in the north and north west, as well as over Colorado, and increased moisture in the south (Figure 5-31). This pattern of drying is consistent with that seen in the background precipitation effects that were just analyzed. In the NVNG simulations for CERA20C and ERA20C there was an average change of -0.00072 and -0.00083 kg/kg in

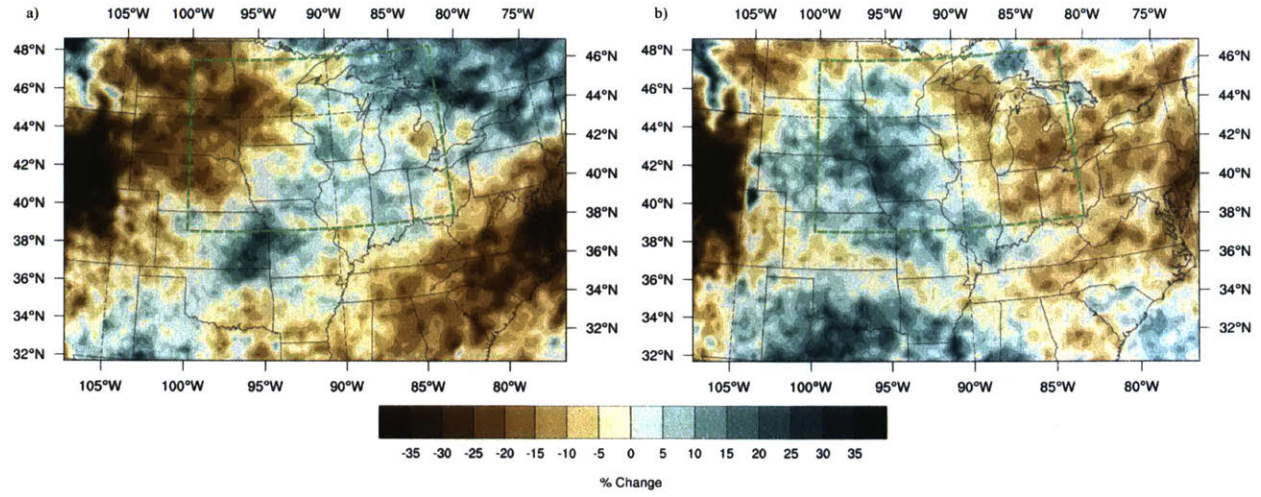


Figure 5-30: July-August average daily precipitation change (%) from 1920-1949 to 1970-1999 in a) CERA20C NVNG b) ERA20C NVNG

the ROSC and -0.00046 and -0.00052 kg/kg in the domain as a whole. These changes were significant in the CERA20C simulation in the domain ($p = 0.00062$) and the ROSC ($p = 0.00062$), as well as in the ERA20C simulation in the same regions ($p = 0.0002$ and $p = 0.000062$).

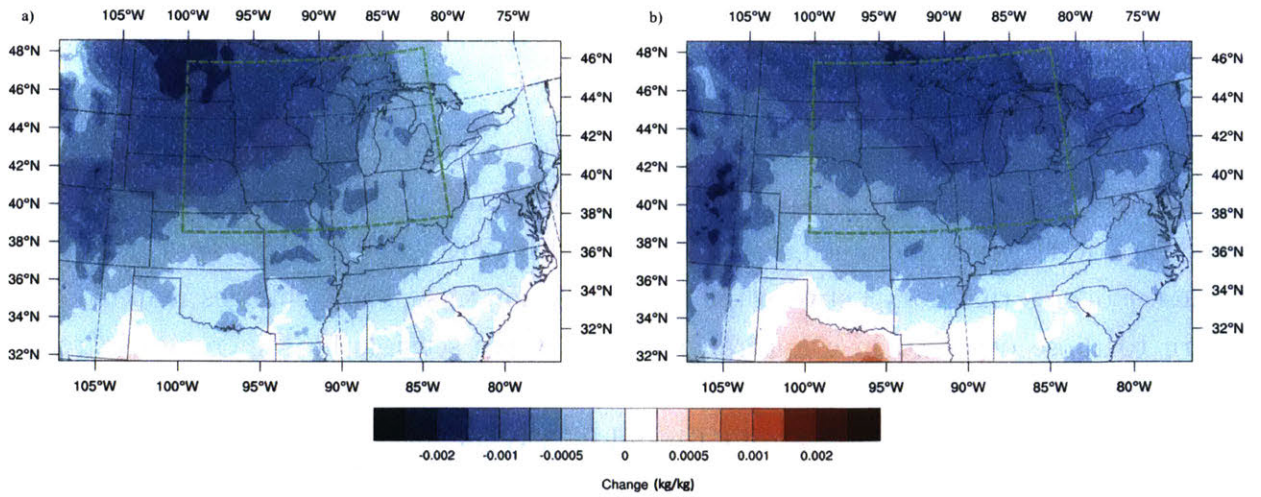


Figure 5-31: July-August average specific humidity change (kg/kg) from 1920-1949 to 1970-1999 in a) CERA20C NVNG b) ERA20C NVNG

The background effects also had a strong impact on evapotranspiration, mostly contributing to a reduction in ET from the early to the late period (Figure 5-32). While this reduction in the strongest in the west, it also encompasses much of the Midwest and particularly irrigated and non-irrigated cropland areas. This influence mutes the overall evapotranspiration

effect in the full VG simulation in both the CERA20C and ERA20C sets and may be a signal that background effects contributed to the underestimation of evapotranspiration that was cited in the comparison to Illinois. In the NVNG simulations for CERA20C and ERA20C there was an average change in evapotranspiration of -0.13 and -0.1 mm/day in the ROSC and -0.08 and -0.09 mm/day in the domain as a whole. These changes were significant in the ERA20C simulation in the domain ($p = 0.00062$) and the ROSC ($p = 0.0017$), and in the CERA20C simulation in the ROSC ($p = 0.0259$) but not in the domain ($p = 0.1088$).

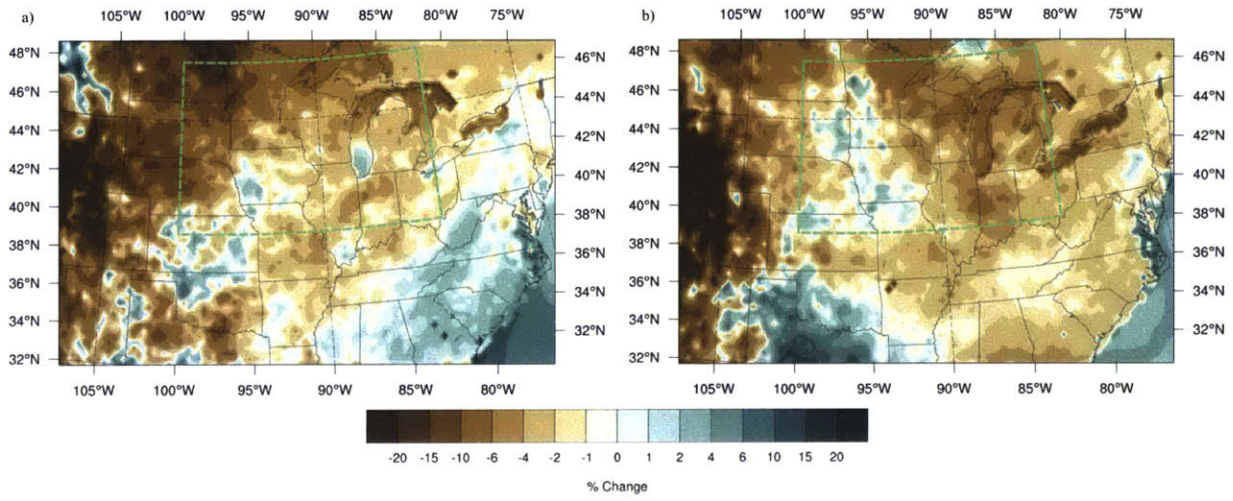


Figure 5-32: July-August average daily evapotranspiration change (%) from 1920-1949 to 1970-1999 in a) CERA20C NVNG b) ERA20C NVNG

Finally, the background effects have a relatively weak impact on the domain sensible heat changes, save for a few hotspots (Figure 5-33). Namely, the NVNG simulations show an increase over Colorado – much more strongly in ERA20C – and some pockets of reduction to a lesser degree that vary between the two simulation sets. These changes map closely in pattern to those observed in evapotranspiration, where sensible heat increases in regions where evapotranspiration decreases and vice versa. For CERA20C and ERA20C there was an average change in sensible heat of 0.61 and 0.93 W/m^2 in the ROSC and 1.16 and 1.66 W/m^2 in the domain. These changes were not significant in the CERA20C simulation in either the ROSC ($p = 0.7600$) or the domain ($p = 0.2003$). They were not significant in the ERA20C simulations in the ROSC ($p = 0.5372$) but were significant in the domain ($p = 0.0113$).

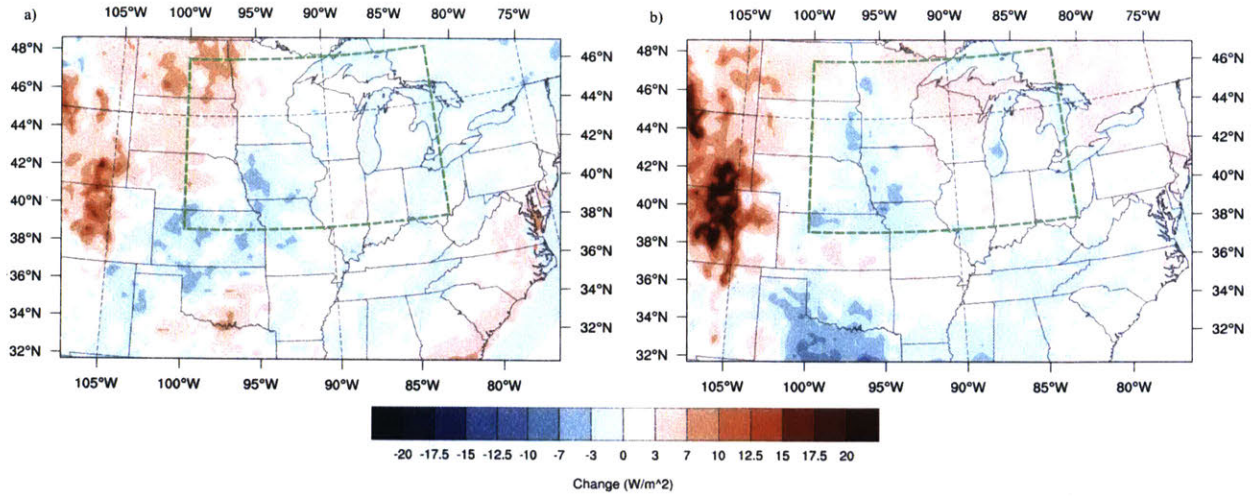


Figure 5-33: July-August average sensible heat change (W/m^2) from 1920-1949 to 1970-1999 in a) CERA20C NVNG b) ERA20C NVNG

5.5.3 Discussion

Because these simulations are a representation of climate changes without the two main forms of anthropogenic forcing represented in the model — greenhouse gases and land use change — the results are due to either internal variability of the climate system, or forcings inherent to the boundary conditions themselves.

Temperature is a good variable to look at to try and understand the forcing of reanalysis data. Both ERA20C and CERA20C reanalysis datasets have available 2m surface temperture data with global coverage and that are available for download and use without any model assimilation beyond production. Additionally, temperature had a strong and well organized impact on the temperature changes in the NVNG simulations.

By looking at the temperature changes present in the reanalysis data, it is evident that the reanalysis data seems to have a strong cold anomaly to the northwest of the domain, in Canada. When mapping the 1920-1949 to 1970-1999 average temperature changes over the entirety of the United States and Canada, this expected cold anomaly is clearly seen in both the CERA20C (Fig 5-34a) and ERA20C (Fig 5-34b) reanalysis whereas it is not present in either CRU (Fig 5-34c) or UDEL (Fig 5-34d).

This analysis leads to the conclusion that the NOAA dataset, despite its coarser resolution may be the better choice for analyzing climate changes over this region and between the two chosen time periods. As a comparison, the same map as in Figure 5-34 is shown for the NOAA 20CR V2c reanalysis is shown in Figure 5-35.

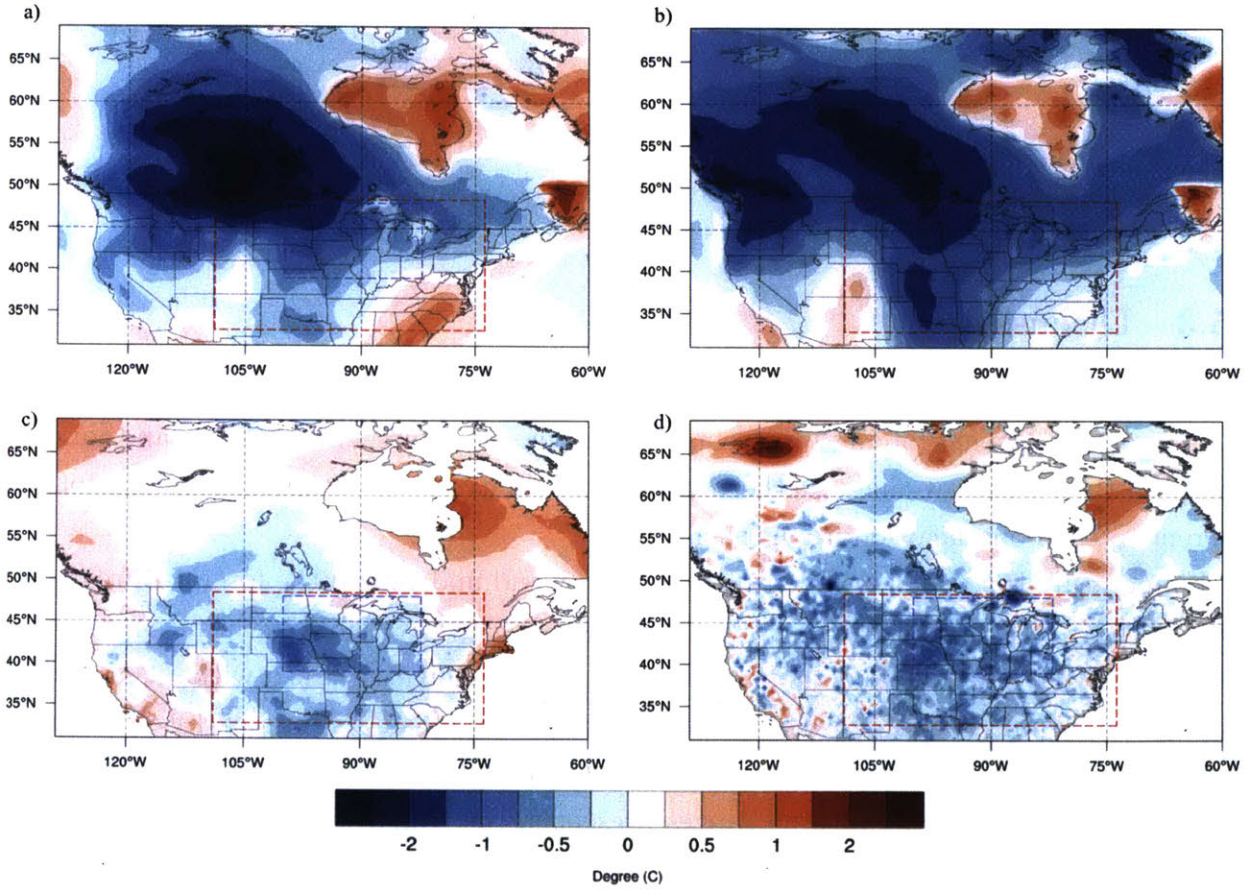


Figure 5-34: July-August average temperature change (°C) from 1920-1949 to 1970-1999 in a) CERA20C Reanalysis b) ERA20C Reanalysis c) CRUTS4.01 and d) UDELv4.01. The red dotted box outlines the domain and the blue dotted box is the ROSC

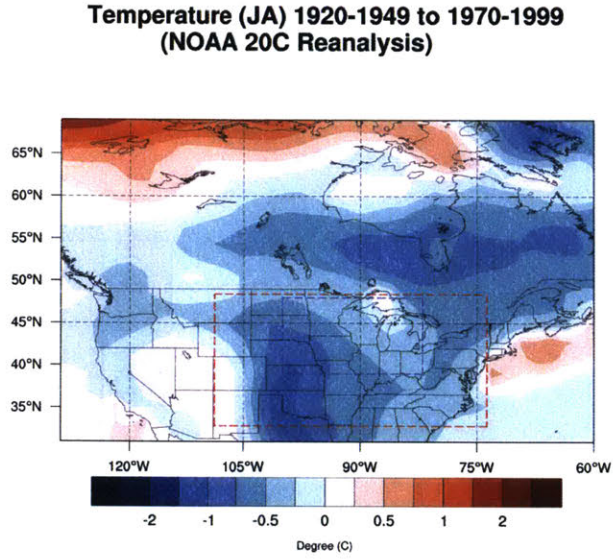
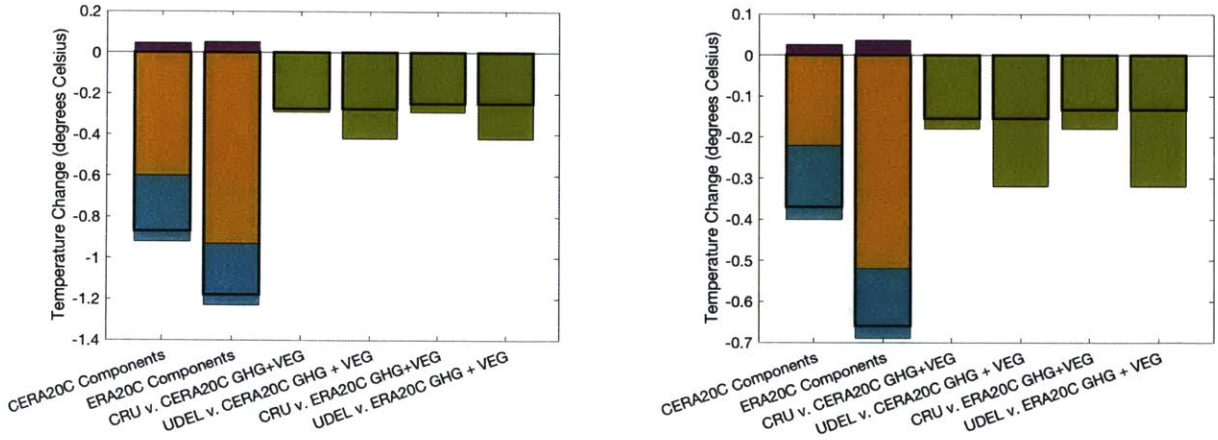


Figure 5-35: July-August average temperature change (°C) from 1920-1949 to 1970-1999 in NOAA 20CR V2c. The red dotted box outlines the domain and the blue dotted box is the ROSC

5.6 Comparison of Factors

As this section has shown, the variables examined are influenced by greenhouse gases, the development of agriculture, and the background effects imposed by reanalysis and internal variability.

These changes can be demonstrated through a series of bar graphs. In each of the graphs the two bars to the far left describe the simulation results for CERA20C and ERA20C respectively. The yellow portion describes the background effects, the purple describes the changes due to GHG, and the blue is the impact due to vegetation. The dark black box in each of these bars shows the composite effect of the three components. To the right of these two bars are comparisons to available observations, shown in green, where the black box denotes the sum of the GHG and vegetation effects for the appropriate simulation for comparison. Where observations are not available, the two bars with model results are still presented to compare the component impacts.

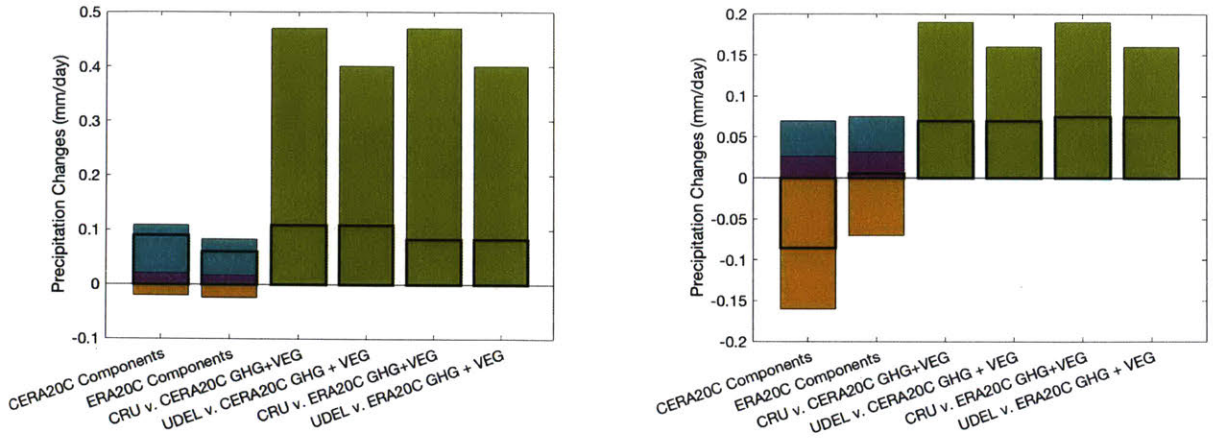


(a) ROSC **(b)** Domain

Figure 5-36: Figures showing selected combinations of temperature changes (°C) for ROSC and the domain.

The background effects composed the majority of the temperature changes between the early to late period as shown by the yellow bars in Figure 5-36. At the same time the vegetation impact was much larger than the greenhouse gas impact in the same direction as the background forcing. The sum of the vegetation and GHG forcings in the model, however, were sufficient to match the observed temperature reduction fairly well, especially in the cases of CRU for both the ROSC and the domain. However, in all cases these combined effects still underestimate the magnitude of the change.

In the case of precipitation, vegetation contributed the most to the change in precipitation



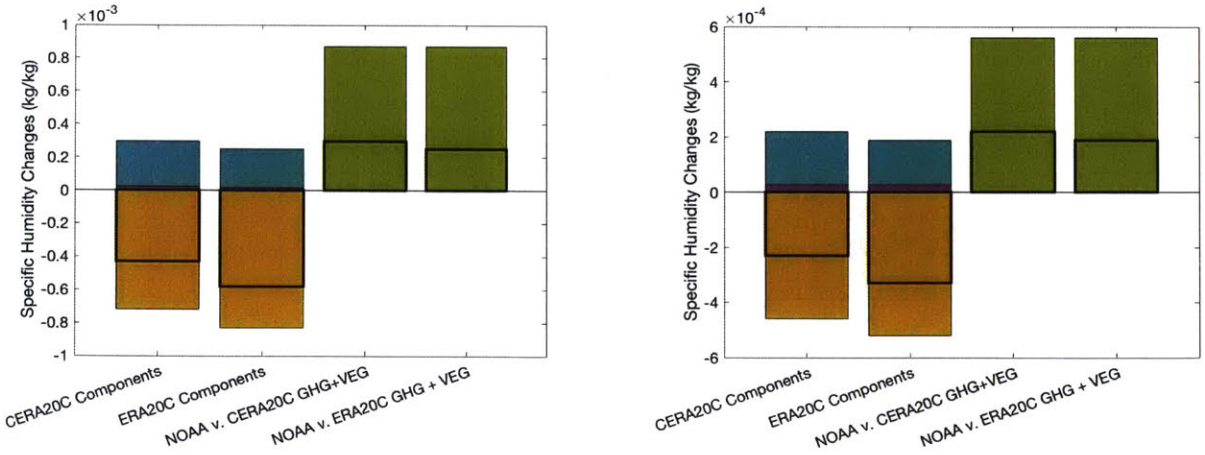
(a) ROSC **(b)** Domain

Figure 5-37: Figures showing selected combinations of daily precipitation changes (mm/day) for ROSC and the domain.

in the ROSC while the background conditions imposed the most change overall in the domain (Figure 5-37). Greenhouse gases were again a smaller component but because of the weak signal overall for precipitation, they played a comparable role towards increasing precipitation in the domain as did vegetation. In this case the sum of vegetation and GHG effects were less able to explain the observed changes, with the combination of these two anthropogenic forcings explaining about a fifth to a fourth of the precipitation changes in the ROSC and a third to nearly half of the change in the domain.

Specific humidity comparisons can be seen in Figure 5-38 and are the final variable for which there are pseudo-observational comparisons between the early and late time period. In this case background effects are again the largest component of change, exceeding the other components by more than double as well as being in the incorrect direction – i.e. showing a decrease in specific humidity. Greenhouse gases play a relatively small role in specific humidity change. The combined impact of GHG and vegetation forcings again underestimates the change indicated in the NOAA dataset, contributing roughly a third of the "observed" change.

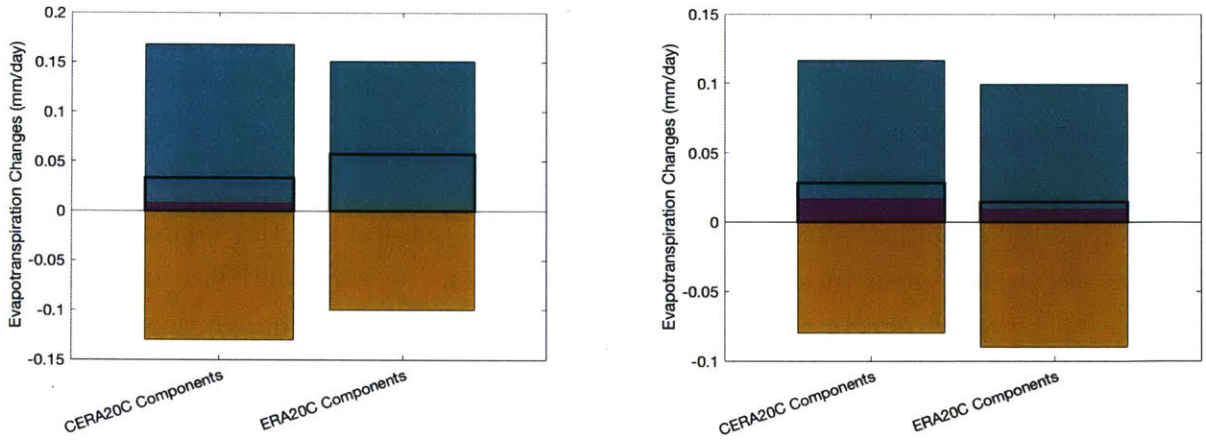
For the last two variables, only the relative components of induced change in daily evapotranspiration (Figure 5-39) and sensible heat flux (Figure 5-40) are considered. For evapotranspiration, in both the ROSC and the domain, vegetation and background effects had comparable magnitudes but work opposite to one another. Greenhouse gases supported the impacts of vegetation but the change was much smaller in proportion to the other two components. For sensible heat flux the same was generally true when looking at the domain, except that vegetation and GHG had opposing effects. However, in the ROSC vegetation impacts



(a) ROSC **(b)** Domain

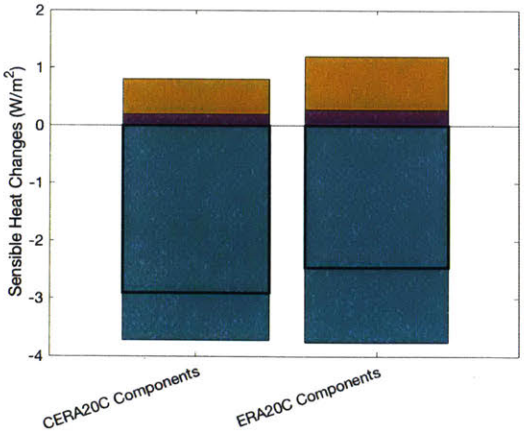
Figure 5-38: Figures showing selected combinations of specific humidity changes (kg/kg) for ROSC and the domain.

were much stronger contributors to the overall change and the effects of the background changes were closer in magnitude to the effect of greenhouse gases.

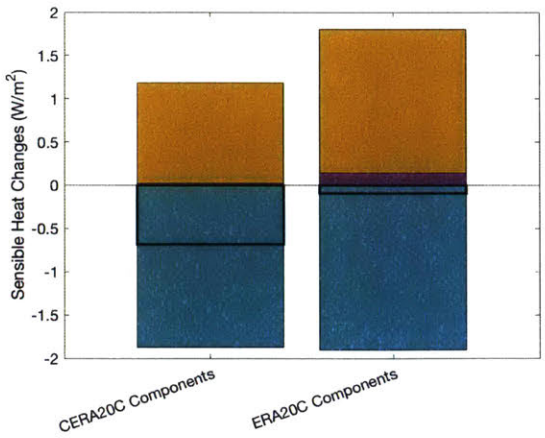


(a) ROSC **(b)** Domain

Figure 5-39: Figures showing selected combinations of daily evapotranspiration changes (mm/day) for ROSC and the domain.



(a) ROSC



(b) Domain

Figure 5-40: Figures showing selected combinations of sensible heat flux changes (W/m^2) for ROSC and the domain.

Overall the greenhouse gas and vegetation developments are having the expected impacts according to the literature in terms of temperature changes, evapotranspiration, sensible heat and specific humidity. The only variable that does not seem to have a strong signal is precipitation, although rainfall is enhanced in the ROSC when looking at vegetation effects relative to GHG or the background effects. Importantly, changes that would be associated with increased plant photosynthesis that have been observed in the observational data seem to be overwhelmed in the model by the strong influence of reanalysis conditions on the domain.

Through these comparisons it is clear that the combination of the two imposed forcings is not adequate to fully explain the changes that occurred in the variables where observational data was available. In no case does the combination of imposed man-made forcings exceed the observed change. This means that the forcings are not being imposed too strongly in the model, but further work is needed to determine the cause of the additionally unexplained variable changes across the century. Either the forcings are more impactful than their current model representation, or the background effects are having a real, sizable impact. Potential avenues for modified representation and component analysis are discussed further in Chapter 6.

Chapter 6

Conclusion

6.1 Summary and Discussion

This study has contributed to climate change research by performing simulations to quantify the relative impacts of greenhouse gases and land use changes on the summer climate of the Midwestern United States in the 20th century. Long term simulations were conducted, rather than performing a simple sensitivity study between two periods. This was made possible by long spanning historical time-series of the forcing variables — cropland, irrigation area, yield, and greenhouse gas concentration — but also by historical reanalysis datasets for the full century that have been developed within the past several years.

The current version of MRCM builds on the experience of several studies performed in the Midwestern United States to improve simulation accuracy. Changes such as the application of irrigation and the enhancement of photosynthesis as detailed in Chapter 4 allow for more realistic representation of land use and its impact on the regional hydrologic cycle.

Based on these simulations, this study looked at the relative contributions of greenhouse gases and land use/land cover change on the observed changes that have occurred in temperature, precipitation and specific humidity. The effects of various forcings have also been explored for evapotranspiration and sensible heat fluxes. The selection of 1920-1949 and 1970-1999 as periods of comparison allowed comparison to the decades that were most likely to characterize the pre- and post-development climate of this area and characterize a period of strong observed change that occurred in the 1950s and 1960s. Additionally, the focus on July-August changes concentrates analysis on the time of high agricultural productivity and matches the time of maximum impact of land use change described in the literature.

Greenhouse gas increases throughout the 20th century have produced a positive change in all of the variables examined in the analysis, although their impact was small in comparison to, and sometimes masked by, the impacts due to vegetation and background effects. On

average, vegetation development caused a reduction in average temperature and sensible heat and an increase in evapotranspiration, specific humidity, and precipitation. These changes are consistent with the expected impacts of productive and irrigated cropland. However, precipitation effects were less cohesive than was expected with the inclusion of irrigated areas. The comparisons for these component effects can be seen in the final section of Chapter 5. According to the simulations conducted here, agricultural development, across all comparisons, had a larger impact on climate change in the region of analysis from 1920-1949 to 1970-1999 than increases in greenhouse gases.

In the case of temperature, the impacts attributed to the combination of greenhouse gases and vegetation development alone were enough to reproduce the observed changes in both the domain and the ROSC, particularly in relation to the CRU data. In the other variables, this combination only explained a portion of the observed changes. In no case where observations were available did the GHG and vegetation components combined exceed the observed change. This shows that the anthropogenic forcings may not have been sufficient to explain observations.

As the above indicates, background effects — those imposed by either internal variability or boundary conditions — resulted in some of the largest effects on the variables, often exceeding and masking the impacts of GHG and vegetation. This study has shown that the choice of reanalysis data is extremely important to the overall magnitude and direction of the final changes, and can occasionally impose a forcing that causes changes either in excess or even opposite to the observed changes.

6.2 Further Questions & Future Work

The study illuminates several focus areas for improvement. As a large portion of model and dataset development was spent on representation of land use and productivity changes, it is important to underscore some assumptions and future work.

6.2.1 Agricultural Representation

Many of the assumptions made in agricultural representation were necessitated by model homogeneity. All crops were considered to be corn, and yield improvements occurred as a single representative step function across all cropland grid cells simultaneously from decade to decade. Several future steps to be taken include:

1. The inclusion of C_3 pathways crops, which encompass soybeans, the other major crop grown in the Midwest. The current model assumption of cropland consisting entirely

of C_4 crops (maize) leads to overestimation of NPP as corn is nearly three times more productive that most other crops. MRCM currently has a distinction between C_3 ("cool") and C_4 ("warm") grasses, with their distribution delineated within the biomes by a temperature threshold. A table showing the defined parameter differences between C_3 and C_4 grasses within the model can be found in table A.1. The introduction of a cropland biome populated by C_3 type crops would allow the designation of a mixture of cropland for each decade. This would allow net primary productivity to be more representative of mixed-cropland. The difficultly in assignment of the biome type would then lie in the lack of explicit crop area distribution records in the pre-remote sensing era, and in the fact that a mixture of crops can be grown within the area of a typical model grid cell.

2. The introduction of an irrigated cropland plant functional type would allow for the specification of a photosynthesis adjustment that could more closely approach the yield increase observed for crops under irrigation (Kucharik & Ramankutty, 2005).

3. The integration of a crop-centric, regionally tuned version of IBIS would allow for more fine scale control on crop growth parameters such as growing season, fertilizer application, and harvest timing (Kucharik, 2003; Kucharik & Brye, 2003; Twine et al., 2004). A more advanced crop model would also more accurately model interannual crop yield in a region that sees the highest interannual variability among corn growing regions in the world (Ben-Ari & Makowski, 2014).

4. In areas such as the Midwest, the introduction of water table dynamics and the explicit representation of groundwater in the model will improve soil moisture results and irrigation dynamics (Yeh & Eltahir, 2005).

Additionally, the results of this study raise several questions related to agricultural representation that can be explored in the future, including: the sensitivity of the model to the application threshold of irrigation for both biome designation and drying level, the impact of allowing crop growth earlier than July, and the component effects of each of the three constituents of agricultural development.

6.2.2 Experiment Inputs and Design

The strong impact of background conditions, and potential issues with the lateral boundary conditions used in the two sets of simulations presented here also suggest improvements that can be made to the model inputs and experimental design.

First, all of the simulations should be repeated with the NOAA 20CR V2c dataset. It is the lowest resolution of the three reanalysis datasets that span the full 20th century, but based on the analysis of surface temperatures shown in Chapter 5, it may impose fewer errors on the simulation domain through the lateral boundaries. The area of strong anomalous cooling in average July-August temperatures present to the northwest of the domain from 1920-1949 to 1970-1999 in ERA20C and CERA20C, is not present in the NOAA data. A comparison of this and the old simulations will also provide the most comprehensive view of the influence of internal variability versus potential reanalysis error. Additionally, future work will involve replicating simulations and providing ensemble-based results, which subdues the impacts of random variability and the chaotic nature of the climate system.

Bibliography

Adegoke, J. O., Pielke Sr., R. A., Eastman, J., Mahmood, R., & Hubbard, K. G. (2003). Impact of irrigation on midsummer surface fluxes and temperature under dry synoptic conditions: A regional atmospheric model study of the U.S. high plains. *Monthly Weather Review*, 131(3), 556–564. https://doi.org/10.1175/1520–0493(2003)131%3C0556:IOIOMS%3E2.0.CO;2

Alter, R. E., Fan, Y., Lintner, B. R., & Weaver, C. P. (2015a). Observational evidence that Great Plains irrigation has enhanced summer precipitation intensity and totals in the midwestern United States. *Journal of Hydrometeorology*, 16(4), 1717–1735. https://doi.org/10.1175/JHM–D–14–0115.1

Alter, R. E., Im, E.-S., & Eltahir, E. A. B. (2015b). Rainfall consistently enhanced around the Gezira scheme in East Africa due to irrigation. *Nature Geoscience Letters*, 8, 763 – 767. https://doi.org/10.1038/NGEO2514

Arakawa, A. & Schubert, W. H. (1974). Interaction of a cumulus cloud ensemble with the large-scale environment, Part I. *Journal of the Atmospheric Sciences*, 31(3), 674–701. https://doi.org/10.1175/1520–0469(1974)031<0674:IOACCE>2.0.CO;2

Ashfaq, M., Rastogi, D., Mei, R., Kao, S.-C., Gangrade, S., Naz, B. S., & Touma, D. (2016). High-resolution ensemble projections of near-term regional climate over the continental United States. *Journal of Geophysical Research: Atmospheres*, 121, 9943–9963. https://doi.org/10.1002/2016JD025653

Bandaru, V., West, T. O., Ricciuto, D. M., & Izaurralde, R. C. (2013). Estimating crop net primary production using national inventory data and MODIS-derived parameters. *ISPRS Journal of Photogrammetry and Remote Sensing*, 80, 61–71. https://10.1016/j.isprsjprs.2013.03.005

Ben-Ari, T. and Makowski, D. (2014). Decomposing global crop yield variability. *Environmental Research Letters*, 9, 114011. https://doi.org/10.1088/1748–9326/9/11/114011

Bhuvandas, N., Timbadiya, P., Patel, P., & Porey, P. (2014). Review of downscaling methods in climate change and their role in hydrological studies. *International Journal of Environmental, Ecological, Geological and Marine Engineering*, 8(10), 648–653.

Bonfils, C. & Lobell, D. (2007). Empirical evidence for a recent slowdown in irrigation-induced cooling. *Proceedings of the National Academy of Sciences of the United States of America*, 104(34), 13582–13587. https://doi.org/10.1073/pnas.0700144104

Bounoua, L., Defries, R., Collatz, G. J., Sellers, P., & Khan, H. (2002). Effects of land cover conversion on surface climate. *Climatic Change*, 52, 29–64. https://doi.org/10.1023/A:1013051420309

Buizza, R., Brönnimann, S., Haimberger, L., Laloyaux, P., Martin, M. J., Fuentes, M., et al. (2017). The EU-FP7 ERA-CLIM2 project contribution to advancing science and production of Earth-system climate reanalyses. *Bulletin of the American Meteorological Society.* Advance online publication. http://journals.ametsoc.org/doi/10.1175/BAMS-D-17-0199.1

Chini, L., Hurtt, G., & Frolking, S. (2014). Harmonized global land use for years 1500 – 2100, v1. Dataset. Available on-line [http://daac.ornl.gov] from Oak Ridge National Laboratory Distributed Active Archive Center, Oak Ridge, Tennessee, USA. http://dx.doi.org/10.3334/ORNLDAAC/1248

Colaizzi, P. D., Lamm, F. R., Howell, T. A., & Evett, S. (2006). Crop production comparison under various irrigation systems. In: Proc. Central Plains Irrigation Conference, Colby, KS., Feb. 21-22, 2006. Available from CPIA. 760 N.Thompson, Colby, KS. pp. 189-207.

Compo, G. P. et al. (2015). NOAA/CIRES Twentieth Century Global Reanalysis Version 2c, https://doi.org/10.5065/D6N877TW, Research Data Archive at the National Center for Atmospheric Research, Computational and Information Systems Laboratory, Boulder, Colo. (Updated yearly)

Compo, G. P., Whitaker, J. S., Sardeshmukh, P. D., Matsui, N., Allan, R. J., Yin, X., et al. (2011). Review article: The twentieth century reanalysis project. *Quarterly Journal of the Royal Meteorological Society*, 137(654), 1–28. https://doi.org/10.1002/qj.776 20th Century Reanalysis V2c data provided by the NOAA/OAR/ESRL PSD, Boulder, Colorado, USA, from their web site at https://www.esrl.noaa.gov/psd/

Cook, B. I., Shukla, S. P., Puma, M. J., & Nazarenko, L. S. (2015). Irrigation as an historical climate forcing. *Climate Dynamics*, 44(5-6), 1715–1730. https://doi.org/10.1007/s00382-014-2204-7

Cunha, A. P. M. A., Alvalá, R. C. S., Sampaio, G., Shimizu, M. H., & Costa, M. H. (2013). Calibration and validation of the Integrated Biosphere Simulator (IBIS) for a Brazilian semiarid region. *Journal of Applied Meteorology and Climatology*, 52, 2753–2770. https://doi.org/10.1175/JAMC-D-12-0190.1

Dai, A. (2012). Increasing drought under global warming in observations and models. *Nature Climate Change*, 3(1), 52–58. https://doi.org/10.1038/nclimate1633

DeAngelis, A., Dominguez, F., Fan, Y., Robock, A., Kustu, M. D., & Robinson, D. (2010). Evidence of enhanced precipitation due to irrigation over the Great Plains of the United States. *Journal of Geophysical Research: Atmospheres*, 115, D15115. https://doi.org/10.1029/2010JD013892

Dell'Aquila, A., Corti, S., Weisheimer, A., Hersbach, H., Peubey, C., Poli, P., Berrisford, P., Dee, D., & Simmons, A. (2016). Benchmarking northern hemisphere midlatitude atmospheric synoptic variability in centennial reanalysis and numerical simulations. *Geophysical Research Letters*, 43(10), 5442–5449. https://doi.org/10.1002/2016GL068829

Diffenbaugh, N. S. (2009). Influence of modern land cover on the climate of the United States. *Climate Dynamics*, 33, 945–958. https://doi.org/10.1007/s00382–009–0566–z

Diffenbaugh, N. S. & Ashfaq, M. (2010). Intensification of hot extremes in the United States. *Geophysical Research Letters*, 37, L15701. https://doi.org/10.1029/2010GL043888

Diffenbaugh, N. S., Pal, J. S., Trapp, R. J., & Giorgi, F. (2005). Fine-scale processes regulate the response of extreme events to global climate change. *Proceedings of the National Academy of Sciences*, 102(44), 15774–15778. https://doi.org/10.1073/pnas.0506042102

Douglas, H. (2016). *Observational analysis of twentieth century summer climate over North America.* (masters thesis), Retrieved from DSpace@MIT. (http://hdl.handle.net/1721.1/104191). Cambridge, MA: Massachusetts Institute of Technology.

Duvick, D. N. (2005). The contribution of breeding to yield advances in maize (Zea mays L.). *Advances in Agronomy*, 86, 83–145. https://doi.org/10.1016/S0065–2113(05)86002–X

ECMWF (2017). Reanalysis Datasets: CERA-20C. Retrieved from https://www.ecmwf.int/en/forecasts/datasets/reanalysis-datasets/cera-20c.

Eltahir, E. A. B. (1998). A soil moisture-rainfall feedback mechanism: 1. Theory and observations. *Water Resources Research*, 34(4), 765–776. https://doi.org/10.1029/97WR03499

ERS (2016). What is agriculture's share of the overall U.S. economy? United States Department of Agriculture, Economic Research Service. Retrieved from https://www.ers.usda.gov/data-products/chart-gallery/gallery/chart-detail/?chartId=58270

FAO (2017). Data. UN Food and Agricultural Organization. Retrieved from http://www.fao.org/faostat/en/data.

FAO, IFAD, & WFP (2015). *The state of food insecurity in the world 2015. Meeting the 2015 international hunger targets: Taking stock of uneven progress.* Rome, FAO.

FAO & WFP (2017). *Monitoring food security in countries with conflict situations.* A joint FAO/WFP update for the United Nations Security Council.

Feitelson, E. & Tubi, A. (2017). A main driver or an intermediate variable? Climate change, water and security in the Middle East. *Global Environmental Change*, 44, 39–48. https://doi.org/10.1016/j.gloenvcha.2017.03.001

Foley, J. A., Prentice, I. C., Ramankutty, N., Levis, S., Pollard, D., Sitch, S., & Haxeltine, A. (1996). An integrated biosphere model of land surface processes, terrestrial carbon balance, and vegetation dynamics. *Global Biogeochemical Cycles*, 10(4), 603–628. https://doi.org/10.1029/96GB02692

Frye, J. D. & Mote, T. L. (2010). Convection initiation along soil moisture boundaries in the southern Great Plains. *Monthly Weather Review*, 138, 1140–1151. https://doi.org/10.1175/2009MWR2865.1

Gianotti, R. L. (2013). *Convective cloud and rainfall processes over the Maritime Continent: Simulation and analysis of the diurnal cycle.* (doctoral dissertation), Retrieved from DSpace@MIT. (http://hdl.handle.net/1721.1/79488). Cambridge, MA: Massachusetts Institute of Technology.

Gianotti, R. L. & Eltahir, E. A. B. (2014a). Regional climate modeling over the Maritime Continent. Part I: New parameterization for convective cloud fraction. *Journal of Climate*, 27(4), 1488–1503. https://doi.org/10.1175/JCLI-D-13-00127.1

Gianotti, R. L. & Eltahir, E. A. B. (2014b). Regional climate modeling over the Maritime Continent. Part II: New parameterization for autoconversion of convective rainfall. *Journal of Climate*, 27(4), 1504–1523. https://doi.org/10.1175/JCLI-D-13-00171.1

Gianotti, R. L., Zhang, D., & Eltahir, E. A. B. (2012). Assessment of the Regional Climate Model version 3 over the Maritime Continent using different cumulus parameterization and land surface schemes. *Journal of Climate*, 25(2), 638–656. https://doi.org/10.1175/JCLI-D-11-00025.1

Glenn, E. P., Huete, A. R., Nagler, P. L., & Nelson, S. G. (2008). Relationship between remotely-sensed vegetation indices, canopy attributes and plant physiological processes: What vegetation indices can and cannot tell us about the landscape. *Sensors*, 8(4), 2136–2160. http://dx.doi.org/10.3390/s8042136

Goldewijk, K. K., Beusen, A., Doelman, J., & Stehfest, E. (2017). Anthropogenic land use estimates for the Holocene - HYDE 3.2. *Earth System Science Data*, 9, 927–953. https://doi.org/10.5194/essd-9-927-2017

Green, J. K., Konings, A. G., Alemohammad, S. H., Berry, J., Entekhabi, D., Kolassa, J., Lee, J.-E., & Gentine, P. (2017). Regionally strong feedbacks between the atmosphere and terrestrial biosphere. *Nature Geoscience*, 10(6), 410–414. https://doi.org/10.1038/ngeo2957

Grell, G. A. (1993). Prognostic evaluation of assumptions used by cumulus parameterizations. *Monthly Weather Review*, 121(3), 764–787. https://doi.org/10.1175/1520-0493(1993)121<0764:PEOAUB>2.0.CO;2

Halder, S., Saha, S. K., Dirmeyer, P. A., Chase, T. N., & Goswami, B. N. (2016). Investigating the impact of land-use land-cover change on Indian summer monsoon daily rainfall and temperature during 1951-2005 using a regional climate model. *Hydrology and Earth System Sciences*, 20, 1765–1784. https://doi.org/10.5194/hess-20-1765-2016

Harding, K. J. & Snyder, P. K. (2012). Modeling the atmospheric response to irrigation in the Great Plains. Part I: General impacts on precipitation and the energy budget. *Journal of Hydrometeorology*, 13, 1667–1686. https://doi.org/10.1175/JHM-D-11-098.1

Harris, I., Jones, P. D., Osborn, T. J., & Lister, D. H. (2014). Updated high-resolution grids of monthly climatic observations - the CRU TS3.10 dataset. *International Journal of Climatology*, 34(3), 623–642. http://doi.org/10/gcmcz3. CRUTS4.01 Data Retrieved From https://crudata.uea.ac.uk/cru/data/hrg/

Heckert, N. & Filliben, J. J. (2003). NIST handbook 148: DATAPLOT reference manual, Volume I: Commands. Technical report, National Institute of Standards and Technology Handbook Series, June 2003, http://www.itl.nist.gov/div898/software/dataplot/refman1/auxillar/ks2samp.htm

Hersbach, H., Peubey, C., Simmons, A., Berrisford, P., Poli, P., & Dee, D. (2015). ERA-20CM: A twentieth-century atmospheric model ensemble. *Quarterly Journal of the Royal Meteorological Society*, 141(691), 2350–2375. https://doi.org/10.1002/qj.2528

Hicke, J. A. & Lobell, D. B. (2004). Spatiotemporal patterns of cropland area and net primary production in the central United States estimated from USDA agricultural information. *Geophysical Research Letters*, 31(20), 1–5. https://doi.org/10.1029/2004GL020927

Hicke, J. A., Lobell, D. B., & Asner, G. P. (2004). Cropland area and net primary production computed from 30 Years of USDA agricultural harvest data. *Earth Interactions*, 8(10), 1–20. https://doi.org/10.1175/1087-3562(2004)008<0001:CAANPP>2.0.CO;2

Hollinger, S. E. & Isard, S. A. (1994). A soil moisture climatology of Illinois. *Journal of Climate*, 7(5), 822–833. https://doi.org/10.1175/1520-0442(1994)007<0822:ASMCOI>2.0.CO;2

Huber, D. B., Mechem, D. B., & Brunsell, N. A. (2014). The effects of Great Plains irrigation on the surface energy balance, regional circulation, and precipitation. *Climate*, 2, 103–128. https://doi.org/10.3390/cli2020103

Hurtt, G. C., Chini, L. P., Frolking, S., Betts, R. A., Feddema, J., Fischer, G., et al. (2011). Harmonization of land-use scenarios for the period 1500-2100: 600 years of global gridded annual land-use transitions, wood harvest, and resulting secondary lands. *Climatic Change*, 109(1), 117–161. https://doi.org/10.1007/s10584-011-0153-2

Im, E.-S. & Eltahir, E. A. B. (2014). Enhancement of rainfall and runoff upstream from irrigation location in a climate model of West Africa. *Water Resources Research*, 50(11), 8651–8674. https://doi.org/10.1002/2014WR015592

Im, E.-S., Marcella, M. P., & Eltahir, E. A. B. (2014). Impact of potential large-scale irrigation on the West African monsoon and its dependence on location of irrigated area. *Journal of Climate*, 27(3), 994–1009. https://doi.org/10.1175/JCLI-D-13-00290.1

IPCC (2014). Climate change 2014: Synthesis report. Contribution of working Groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. [Core Writing Team, R.K. Pachauri and L.A. Meyer (eds.)], Geneva, Switzerland: IPCC. 151 pp.

Jaafar, H. H. & Ahmad, F. A. (2015). Crop yield prediction from remotely sensed vegetation indices and primary productivity in arid and semi-arid lands. *International Journal of Remote Sensing*, 36(18), 4570–4589. http://dx.doi.org/10.1080/01431161.2015.1084434

Jin, J. & Miller, N. L. (2011). Regional simulations to quantify land use change and irrigation impacts on hydroclimate in the California Central Valley. *Theoretical Applied Climatology*, 104, 429–442. https://doi.org/10.1007/s00704-010-0352-1

Koster, R. D., Dirmeyer, P. A., Guo, Z., Bonan, G., Chan, E., Cox, P., and et al. (2004). Regions of strong coupling between soil moisture and precipitation. *Science*, 305, 1138–1140. https://doi.org/10.1126/science.1100217

Kucharik, C. J. (2003). Evaluation of a process-based agro-ecosystem model (Agro-IBIS) across the U.S. corn belt: Simulations of the interannual variability in maize yield. *Earth Interactions*, 7(14), 1–33. https://doi.org/10.1175/1087-3562(2003)007<0001:EOAPAM>2.0.CO;2

Kucharik, C. J. & Brye, K. R. (2003). Integrated BIosphere Simulator (IBIS) yield and nitrate loss predictions for Wisconsin maize receiving varied amounts of nitrogen fertilizer. *Journal of Environmental Quality*, 32, 247–268. https://doi.org/10.2134/jeq2003.2470

Kucharik, C. J., Foley, J. A., Delire, C., Fisher, V. A., Coe, M. T., Lenters, J. D., et al. (2000). Testing the performance of a dynamic global ecosystem model: Water balance, carbon balance, and vegetation structure. *Global Biogeochemical Cycles*, 14(3), 795–825. http://dx.doi.org/10.1029/1999GB001138

Kucharik, C. J. & Ramankutty, N. (2005). Trends and variability in U.S. corn yields over the twentieth century. *Earth Interactions*, 9(1), 1–29. https://doi.org/10.1175/EI098.1

Kueppers, L. M., Snyder, M. A., & Sloan, L. C. (2007). Irrigation cooling effect: Regional climate forcing by land-use change. *Geophysical Research Letters*, 34, L03703. https://doi.org/10.1029/2006GL028679

Laloyaux, P. (2017). The Climate Data Guide: CERA-20C: ECMWF's Coupled Ocean-Atmosphere Reanalysis of the 20th Century. National Center for Atmospheric Research Staff (Eds.). Retrieved from https://climatedataguide.ucar.edu/climate-data/cera-20c-ecmwfs-coupled-ocean-atmosphere-reanalysis-20th-century.

Laloyaux, P., Balmaseda, M., Dee, D., Mogensen, K., & Janssen, P. (2016). A coupled data assimilation system for climate reanalysis. *Quarterly Journal of the Royal Meteorological Society*, 142(694), 65–78. https://doi.org/10.1002/qj.2629

Levis, S., Coe, M. T., & Foley, J. A. (1996). Hydrologic budget of a land surface model: A global application. *Journal of Geophysical Research*, 101(D12), 16921–16930. https://doi.org/10.1029/96JD01164

Li, Z., Liu, S., Tan, Z., Bliss, N. B., Young, C. J., West, T. O., & Ogle, S. M. (2014). Comparing cropland net primary production estimates from inventory, a satellite-based model, and a process-based model in the Midwest of the United States. *Ecological Modelling*, 277, 1–12. https://doi.org/10.1016/j.ecolmodel.2014.01.012

Lo, M.-H. & Famiglietti, J. S. (2013). Irrigation in California's Central Valley strengthens the southwestern U.S. water cycle. *Geophysical Research Letters*, 40(2), 301–306. https://doi.org/10.1002/grl.50108

Lobell, D. B. & Asner, G. P. (2003). Climate and management contributions to recent trends in U.S. agricultural yields. *Science*, 299, 1032. https://doi.org/10.1126/science.1078475

Lobell, D. B., Bonfils, C. J., Kueppers, L. M., & Snyder, M. A. (2008). Irrigation cooling effect on temperature and heat index extremes. *Geophysical Research Letters*, 35(9), L09705. https://doi.org/10.1029/2008GL034145

Lobell, D. B., Hicke, J. A., Asner, G. P., Field, C. B., Tucker, C. ., & Loss, S. O. (2002). Satelite estimates of productivity and light use efficiency in United States agriculture, 1982-1998. *Global Change Biology*, 8, 722–735. https://doi.org/10.1046/j.1365–2486.2002.00503.x

Lobell, D. B., Roberts, M. J., Schlenker, W., Braun, N., Little, B. B., Rejesus, R. M., & Hammer, G. L. (2014). Greater sensitivity to drought accompanies maize yield increase in the U.S. Midwest. *Science*, 344, 516–519. https://doi.org/10.1126/science.1251423

Long, S. P., Zhu, X.-G., Naidu, S. L., & Ort, D. R. (2006). Can improvement in photosynthesis increase crop yields? *Plant, Cell and Environment*, 29(3), 315–330. https://doi.org/10.1111/j.1365–3040.2005.01493.x

Lu, Y., Harding, K., & Kueppers, L. M. (2017). Irrigation effects on land-atmosphere coupling strength in the United States. *Journal of Climate*, 30(10), 3671–3685. https://doi.org/10.1175/JCLI–D–15–0706.1

Malyshev, S., Shevliakova, E., Stouffer, R. J., & Pacala, S. W. (2015). Contrasting local versus regional effects of land-use-change-induced heterogeneity on historical climate: Analysis with the GFDL earth system model. *Journal of Climate*, 28(13), 5448–5469. https://doi.org/10.1175/JCLI–D–14–00586.1

Marcella, M. P. (2013). *Biosphere-atmosphere interactions over semi-arid regions: Modeling the role of mineral aerosols and irrigation in the regional climate system.* (doctoral dissertation), Retrieved from DSpace@MIT. (http://hdl.handle.net/1721.1/79490). Cambridge, MA: Massachusetts Institute of Technology.

Marcella, M. P. & Eltahir, E. A. B. (2010). Effects of mineral aerosols on the summertime climate of southwest Asia: Incorporating subgrid variability in a dust emission scheme. *Journal of Geophysical Research*, 115, D18203. https://doi.org/10.1029/2010JD014036

Marcella, M. P. & Eltahir, E. A. B. (2012). Modeling the summertime climate of Southwest Asia: The role of land surface processes in shaping the climate of semiarid regions. *Journal of Climate*, 25, 704–719. https://doi.org/10.1175/2011JCLI4080.1

Maupin, M. A., Kenny, J. F., Hutson, S. S., Lovelace, J. K., Barber, N. L., & Linsey, K. S. (2014). Estimated use of water in the United States in 2010: U.S. Geological Survey circular 1405, 56 p. http://dx.doi.org/10.3133/cir1405

Meinshausen, M., Smith, S. J., Calvin, K., Daniel, J. S., Kainuma, M. L., Lamarque, J.-F., et al. (2011). The RCP greenhouse gas concentrations and their extensions from 1765 to 2300. *Climatic Change*, 109, 213–241. https://doi.org/10.1007/s10584–011–0156–z. Data Retrieved from http://www.pik–potsdam.de/ mmalte/rcps/

Melillo, J. M., Richmod, T. C., & Yohe, G. W. (2014). Climate change impacts in the United States: The third national climate assessment. U.S. Global Change Research Program, 841 pp. https://doi.org/10.7930/J0Z31WJ2

Mishra, V., Cherkauer, K. A., & Shukla, S. (2010). Assessment of drought due to historic climate variability and projected future climate change in the Midwestern United States. *Journal of Hydrometeorology*, 11(1), 46–68. https://doi.org/10.1175/2009JHM1156.1

Monfreda, C., Ramankutty, N., & Foley, J. A. (2008). Farming the planet: 2. Geographic distribution of crop areas, yields, physiological types, and net primary production in the year 2000. *Global Biogeochemical Cycles*, 22, GB1022. http://dx.doi.org/10.1029/2007GB002947

Mueller, N. D., Butler, E. E., McKinnon, K. A., Rhines, A., Tingley, M., Holbrook, N. M., & Huybers, P. (2016). Cooling of U.S. Midwest summer temperature extremes from cropland intensification. *Nature Climate Change*, 6(3), 317–322. http://dx.doi.org/10.1038/nclimate2825

NCAR Staff (Eds.) (2017). The Climate Data Guide: NOAA 20th-Century Reanalysis, Version 2 and 2c. Retrieved from https://climatedataguide.ucar.edu/climate-data/noaa-20th-century-reanalysis-version-2-and-2c.

Nickerson, C. and Borchers, A. (2012). How is land in the United States used? a focus on agricultural land. United States Department of Agriculture, Economic Research Service, Amber Waves. https://www.ers.usda.gov/amber-waves/2012/march/data-feature-how-is-land-used/

Nordås, R. & Gleditsch, N. P. (2007). Climate change and conflict. *Political Geography*, 26(6), 627–638. https://doi.org/10.1016/j.polgeo.2007.06.003

Ozdogan, M., Rodell, M., Beaudoing, H. K., & Toll, D. L. (2010). Simulating the effects of irrigation over the United States in a land surface model based on satellite-derived agricultural data. *Journal of Hydrometeorology*, 11, 171–184. https://doi.org/10.1175/2009JHM1116.1

Pal, J. S. & Eltahir, E. A. B. (2002). Teleconnections of soil moisture and rainfall during the 1993 Midwest summer flood. *Geophysical Research Letters*, 29(18), 12–1 - 12–4. https://doi.org/10.1029/2002GL014815

Peixoto, J. P. & Oort, A. H. (1992). *Physics of Climate*. New York, NY: Springer-Verlag.

Peterson, T. C. and Vose, R. S. (1997). An overview of the Global Historical Climatology Network temperature database. *Bulletin of the American Meteorological Society*, 78(12), 2837–2849. https://doi.org/10.1175/1520–0477(1997)078<2837:AOOTGH>2.0.CO;2

Poli, P., Hersbach, H., Dee, D. P., Berrisford, P., Simmons, A. J., Vitart, F., et al. (2016). ERA-20C: An atmospheric reanalysis of the twentieth century. *Journal of Climate*, 29(11), 4083–4097. https://doi.org/10.1175/JCLI–D–15–0556.1

Poli, P., Hersbach, H., Dee, D. P., Thépaut, J. N., Simmons, A. J., Peubey, C., et al. (2013). The data assimilation system and initial performance evaluation of the ECMWF pilot reanalysis of the 20th-century assimilating surface observations only (ERA-20C). Technical Report 14, European Centre for Medium-Range Weather Forecasts (ECMWF), ERA Report Series.

Pollard, D. & Thompson, S. L. (1995). Use of a land-surface-transfer scheme (LSX) in a global climate model: the response to doubling stomatal resistance. *Global and Planetary Change*, 10, 129–161. https://doi.org/10.1016/0921–8181(94)00023–7

Prince, S. D., Haskett, J., Steininger, M., Strand, H., & Wright, R. (2001). Net primary production of U.S. Midwest croplands from agricultural harvest yield data. *Ecological Applications*, 11(4), 1194–1205. http://dx.doi.org/10.1890/1051–0761(2001)011[1194:NPPOUS]2.0.CO;2

Qian, Y., Huang, M., Yang, B., & Berg, L. K. (2013). A modeling study of irrigation effects on surface fluxes and land-air-cloud interactions in the southern Great Plains. *Journal of Hydrometeorology*, 14, 700–721. https://doi.org/10.1175/JHM–D–12–0134.1

Ramankutty, N. & Foley, J. A. (1999). Estimating historical changes in global land cover: Croplands from 1700 to 1992. *Global Biogeochemical Cycles*, 13(4), 997–1027. https://doi.org/10.1029/1999GB900046

Ray, D. K., Gerber, J. S., MacDonald, G. K., & West, P. C. (2015). Climate variation explains a third of global crop yield variability. *Nature Communications*, 6(5989). http://doi.org/10.1038/ncomms6989

Ray, D. K., Mueller, N. D., West, P. C., & Foley, J. A. (2013). Yield trends Are insufficient to double global crop production by 2050. *PLoS ONE*, 8(6), e66248. https://doi.org/10.1371/journal.pone.0066428

Ray, D. K., Ramankutty, N., Mueller, N. D., West, P. C., & Foley, J. A. (2012). Recent patterns of crop yield growth and stagnation. *Nature Communications*, 3(1293). http://www.nature.com/doifinder/10.1038/ncomms2296

Rayner, N. A., Parker, D. E., Horton, E. B., Folland, C. K., Alexander, L. V., Rowell, D. P., Kent, E. C., & Kaplan, A. (2003). Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century. *Journal of Geophysical Research*, 108(D14). https://doi.org/10.1029/2002JD002670. Data Retrieved from https://www.metoffice.gov.uk/hadobs/hadisst/data/download.html

Reeves, M. C., Zhao, M., & Running, S. W. (2005). Usefulness and limits of MODIS GPP for estimating wheat yield. *International Journal of Remote Sensing*, 26(7), 1403–1421. http://dx.doi.org/10.1080/01431160512331326567

Reuveny, R. (2007). Climate change-induced migration and violent conflict. *Political Geography*, 26(6), 656–673. https://doi.org/10.1016/j.polgeo.2007.05.001

Rosenzweig, M. L. (1968). Net primary productivity of terrestrial communities: Prediction from climatological data. *The American Naturalist*, 102(923), 67–74. https://doi.org/10.1086/282523

Roser, M. & Ortiz-Ospina, E. (2018). "World Population Growth". *Published online at OurWorldInData.org.* Retrieved from https://ourworldindata.org/world-population-growth.

Roser, M. & Ritchie, H. (2018). "Yields and Land Use in Agriculture". *Published online at OurWorldInData.org.* Retrieved from https://ourworldindata.org/yields-and-land-use-in-agriculture.

Roxburgh, S. H., Berry, S. L., Buckley, T. N., Barnes, B., & Roderick, M. L. (2005). What is NPP? Inconsistent accounting of respiratory fluxes in the definition of net primary production. *Functional Ecology*, 19(3), 378–382. https://doi.org/10.1111/j.1365–2435.2005.00983.x

Sacks, W. J. & Kucharik, C. J. (2011). Crop management and phenology trends in the U.S. Corn Belt: Impacts on yields, evapotranspiration and energy balance. *Agricultural and Forest Meteorology*, 151(7), 882–894. http://dx.doi.org/10.1016/j.agrformet.2011.02.010

Seneviratne, S. I., Corti, T., Davin, E. L., Hirschi, M., Jaeger, E. B., Lehner, I., Orlowsky, B., & Teuling, A. J. (2010). Investigating soil moisture-climate interactions in a changing climate: A review. *Earth-Science Reviews*, 99, 125–161. https://doi.org/10.1016/j.earscirev.2010.02.004

Sheskin, D. J. (2007). *Handbook of parametric and nonparametric statistical procedures.* (Fourth Ed.). Boca Raton, FL: Chapman Hall/CRC.

Siebert, S., Kummu, M., Porkka, M., Döll, P., Ramankutty, N., & Scanlon, B. R. (2015). A global data set of the extent of irrigated land from 1900 to 2005. *Hydrology and Earth System Sciences*, 19, 1521–1545. Data Available at http://dx.doi.org/10.13019/M20599

SRB Science Team (2010). Surface Radiation Budget (SRB) Release 3.0 GEWEX Shortwave Monthly UTC Data, Hampton, VA, USA: NASA Atmospheric Science Data Center (ASDC), Retrieved from https://doi.org/10.5067/SRB/REL3.0_SW_MONTHLY_UTC_NC_L3

SRB Science Team (2012). Surface Radiation Budget (SRB) Release 3.1 GEWEX Longwave monthly Data, Hampton, VA, USA: NASA Atmospheric Science Data Center (ASDC), Retrieved from https://doi.org/10.5067/SRB/REL3.1_LW_MONTHLY_NC_L2

Sterling, S. M., Ducharne, A., & Polcher, J. (2012). The impact of global land-cover change on the terrestrial water cycle. *Nature Climate Change*, 3(4), 385–390. https://doi.org/10.1038/nclimate1690

Stickler, A., Storz, S., Wartenburger, R., Hersbach, H., Compo, G. P., Poli, P., Dee, D., & Bronnimann, S. (2015). Upper-air observations from the German Atlantic expedition (1925-1927) and comparison with the twentieth century and ERA-20C reanalyses. *Meteorologische Zeitschrift*, 24(5), 525–544. https://doi.org/10.1127/metz/2015/0683

Stryhal, J. & Huth, R. (2017). Classifications of winter Euro-Atlantic circulation patterns: An intercomparison of five atmospheric reanalyses. *Journal of Climate*, 30(19), 7847–7861. https://doi.org/10.1175/JCLI-D-17-0059.1

Tollenaar, M. & Lee, E. A. (2006). Dissection of physiological processes underlying grain yield in maize by examining genetic improvement and heterosis. *Maydica*, 51, 399–408.

Turner, D. P., Ritts, W. D., Cohen, W. B., Gower, S. T., Running, S. W., Zhao, M., et al. (2006). Evaluation of MODIS NPP and GPP products across multiple biomes. *Remote Sensing of Environment*, 102(3-4), 282–292. https://doi.org/10.1016/j.rse.2006.02.017

Twine, T. E., Bryant, J. J., Richter, K. T., Bernacchi, C. J., Mcconnaughay, K. D., Morris, S. J., & Leakey, A. D. B. (2013). Impacts of elevated CO2 concentration on the productivity and surface energy budget of the soybean and maize agroecosystem in the Midwest USA. *Global Change Biology*, 19(9), 2838–2852. https://doi.org/10.1111/gcb.12270

Twine, T. E. & Kucharik, C. J. (2009). Climate impacts on net primary productivity trends in natural and managed ecosystems of the central and eastern United States. *Agricultural and Forest Meteorology*, 149(12), 2143–2161. https://doi.org/10.1016/j.agrformet.2009.05.012

Twine, T. E., Kucharik, C. J., & Foley, J. A. (2004). Effects of land cover change on the energy and water balance of the Mississippi River basin. *Journal of Hydrometeorology*, 5(1982), 640–655. https://doi.org/10.1175/1525–7541(2004)005%3C0640:EOLCCO%3E2.0.CO;2

UN (2017). *World Population Prospects: The 2017 Revision, Key Findings and Advance Tables.* United Nations, Department of Economic and Social Affairs, Population Division. Working Paper No. ESA/P/WP/248.

USDA (1999). *Weekly Weather and Crop Bulletin*, volume 86. No. 18, 22, 27, 31, 36, 40, 44. U.S. Department of Agriculture.

USDA (2014). 2012 census of agriculture: Farm and ranch irrigation survery (2013) (volume 3, special studies, part 1). U.S. Department of Agriculture National Agricultural Statistics Service.

USDA (2016). *Weekly Weather and Crop Bulletin*, volume 103. No. 18, 23, 27, 31, 36, 40, 45. U.S. Department of Agriculture.

USDA-NASS (2016). Quick Stats. U.S. Department of Agriculture National Agricultural Statisitics Service (USDA NASS). Retrieved from https://quickstats.nass.usda.gov/.

West, T. O., Brandt, C. C., Baskaran, L. M., Hellwinckel, C. M., Mueller, R., Bernacchi, C. J., et al. (2010). Cropland carbon fluxes in the United States: Increasing geospatial resolution of inventory-based carbon accounting. *Ecological Applications*, 20(4), 1074–1086. https://doi.org/10.1890/08–2352.1

Willmott, C. J. & Matsuura, K. (2015a). Terrestrial air temperature: 1900-2014 gridded monthly time series (1900-2014). `http://climate.geog.udel.edu/~climate/html_pages/Global2014/README.GlobalTsT2014.html`.

Willmott, C. J. & Matsuura, K. (2015b). Terrestrial Precipitation: 1900-2014 Gridded Monthly Time Series (1900-2014). `http://climate.geog.udel.edu/~climate/html_pages/Global2014/README.GlobalTsP2014.html`.

Winter, J. M. (2006). *Coupling of Integrated BIosphere Simulator to Regional Climate Model version 3.* (masters thesis), Retrieved from DSpace@MIT. (http://hdl.handle.net/1721.1/34272). Cambridge, MA: Massachusetts Institute of Technology.

Winter, J. M. (2010). *Modeling land surface processes of the Midwestern United States: Predicting soil moisture under a warmer climate.* (doctoral dissertation), Retrieved from DSpace@MIT. (http://hdl.handle.net/1721.1/60707). Cambridge, MA: Massachusetts Institute of Technology.

Winter, J. M. & Eltahir, E. A. B. (2010). The sensitivity of latent heat flux to changes in the radiative forcing: A framework for comparing models and observations. *Journal of Climate*, 23, 2345–2356. https://doi.org/10.1175/2009JCLI3158.1

Winter, J. M. & Eltahir, E. A. B. (2012a). Modeling the hydroclimatology of the Midwestern United States. Part 1: Current climate. *Climate Dynamics*, 38, 573–593. https://doi.org/10.1007/s00382–011–1182–2

Winter, J. M. & Eltahir, E. A. B. (2012b). Modeling the hydroclimatology of the Midwestern United States. Part 2: Future climate. *Climate Dynamics*, 38, 595–611. https://doi.org/10.1007/s00382–011–1183–1

Winter, J. M., Yeh, P. J.-F., Fu, X., & Eltahir, E. A. B. (2015). Uncertainty in modeled and observed climate change impacts on American Midwest hydrology. *Water Resources Research*, 50, 1649–1670. https://doi.org/10.1002/2015WR017096

Wood, E. F. (1997). Effects of soil moisture aggregation on surface evaporative fluxes. *Journal of Hydrology*, 190, 397–412. https://doi.org/10.1016/S0022-1694(96)03135-6

World Bank (2017). World Development Indicators. Retrieved from https://data.worldbank.org/indicator.

Xin, Q., Broich, M., Suyker, A. E., Yu, L., & Gong, P. (2015). Multi-scale evaluation of light use efficiency in MODIS gross primary productivity for croplands in the Midwestern United States. *Agricultural and Forest Meteorology*, 201, 111–119. https://doi.org/10.1016/j.agrformet.2014.11.004

Yeh, P. J.-F. & Eltahir, E. A. B. (2005). Representation of water table dynamics in a land surface scheme. Part I: Model development. *Journal of Climate*, 18(12), 1861–1880. https://doi.org/10.1175/JCLI3330.1

Yeh, P. J.-F., Irizarry, M., & Eltahir, E. A. B. (1998). Hydroclimatology of Illinois: A comparison of monthly evaporation estimates based on atmospheric water balance and soil water balance. *Journal of Geophysical Research*, 103(D16), 19823–19837. https://doi.org/10.1029/98JD01721

Zhao, M., Heinsch, F. A., Nemani, R. R., & Running, S. W. (2005). Improvements of the MODIS terrestrial gross and net primary production global data set. *Remote Sensing of Environment*, 95, 164–176. https://10.1016/j.rse.2004.12.011

Appendix A

Tables

Table A.1: Defined parameter differences between cool (C3) and warm (C4) grasses within IBIS (Foley et al., 1996)

Parameter	C3 variable name and value	C4 variable name and value
Intrinsic Quantum Efficiency	alpha3 = 0.060	alpha4 = 0.050
Leaf Respiration Coefficients	gammal3 = 0.0150	gammal4 = 0.0300
'm' Coeff. for Stomatal Conductance	coefml3 = 9.0	coefml4 = 4.0
'b' Coeff. for Stomatal Conductance	coeffbl3 = 0.010	coeffbl4 = 0.040
Absolute Min. Stomatal Conductance	gsl3min = 0.00001	gsl4min = 0.00001
Photosynthesis Coupling Coeff.	theta3 = 0.950 beta3 = 0.990	theta4 = 0.970 beta4 = 0.800
Max. Rubsico Activity (Top of Canopy)	vmaxl3 = 2.0e-06	vmaxl4 = 15.0e-06

Appendix B

Figures

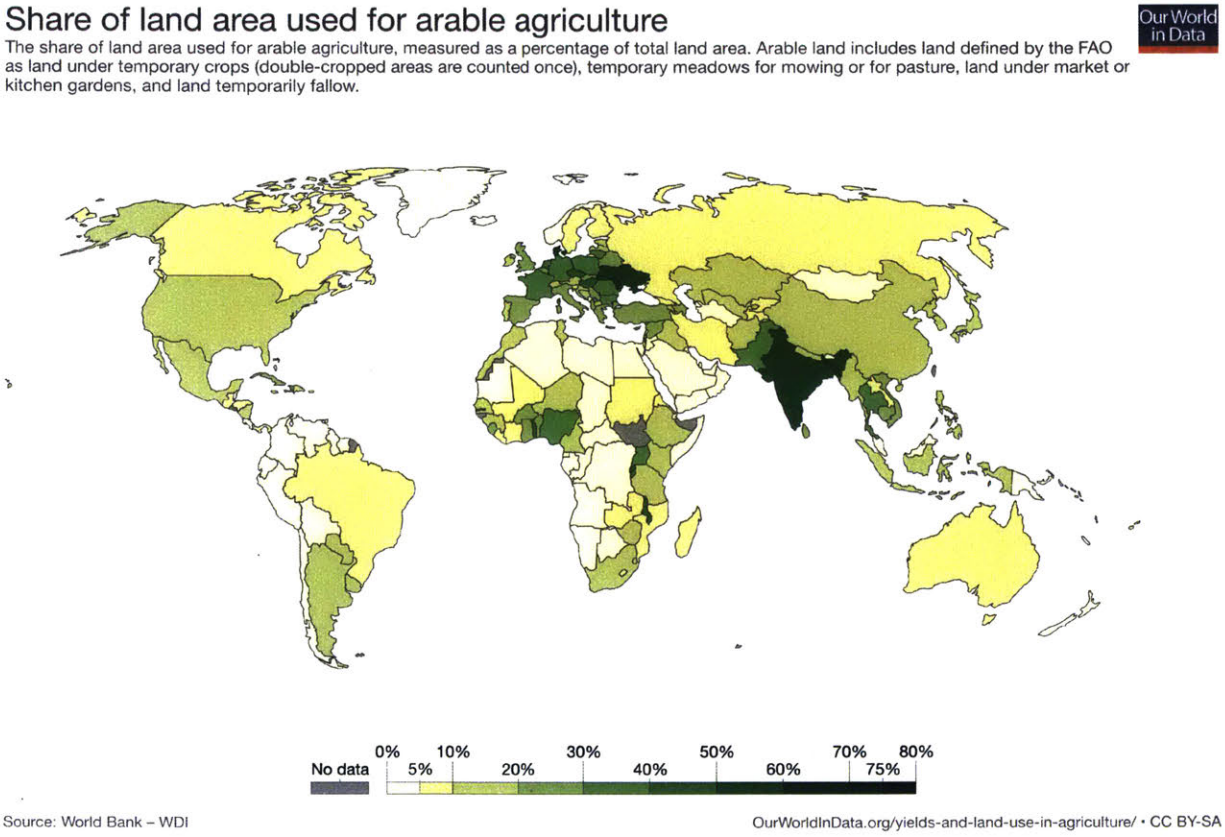


Figure B-1: Share of land area used for arable agriculture

Figure available at https://ourworldindata.org/grapher/share-of-land-area-used-for-arable-agriculture?tab=chart (Data Source: World Bank, 2017)

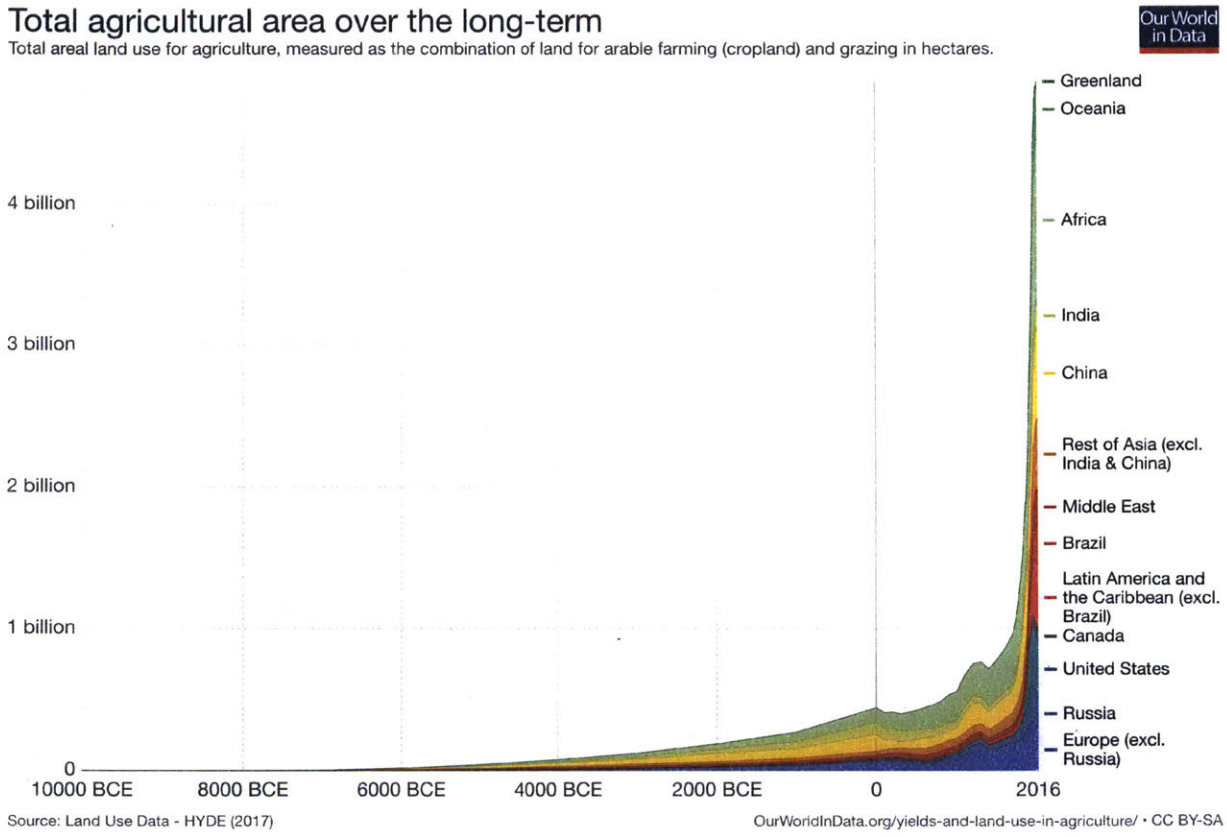


Figure B-2: Total agricultural area over the long-term

Figure available at https://ourworldindata.org/grapher/total-agricultural-area-over-the-long-term (Data Source: Goldewijk et al., 2017)

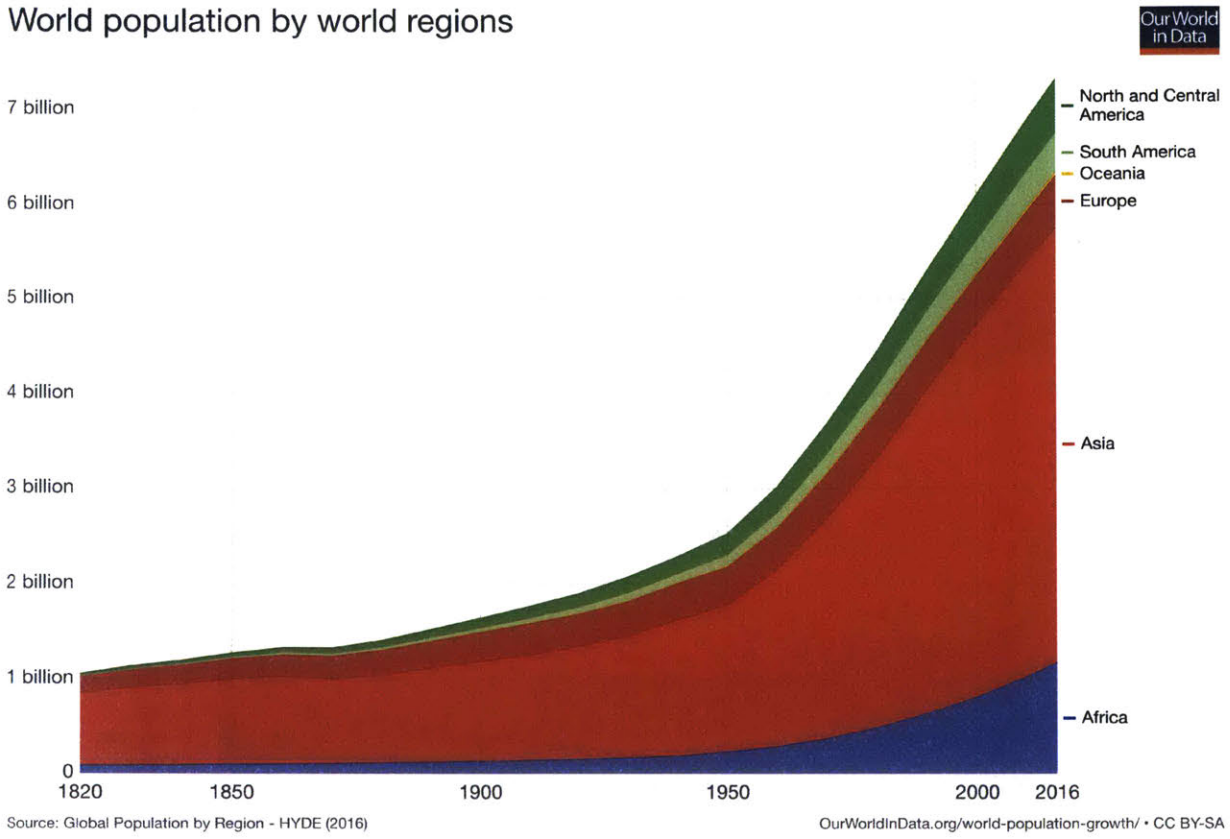


Figure B-3: World population by world regions

Figure avialable at https://ourworldindata.org/grapher/world-population-by-world-regions-post-1820. (Data Source: Goldewijk et al., 2017)

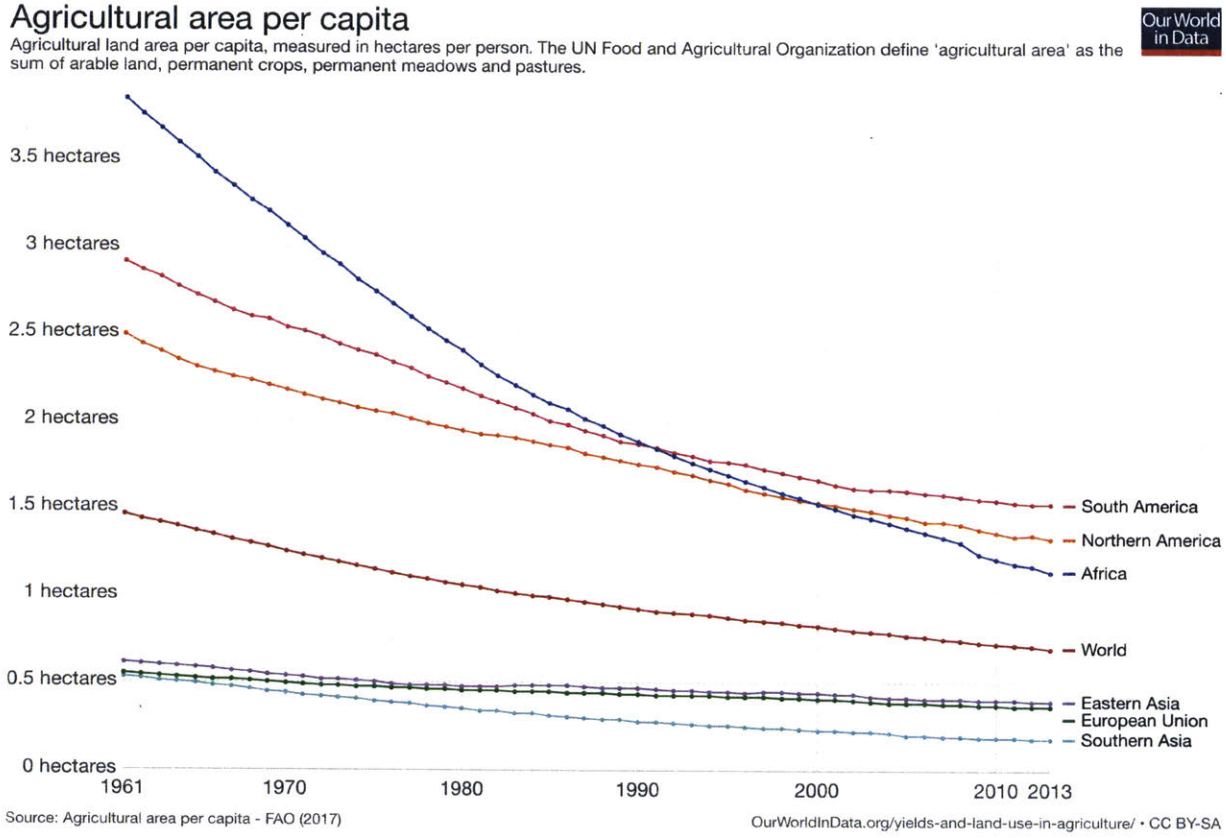


Figure B-4: Agricultural area per capita

Figure available at https://ourworldindata.org/grapher/agricultural-area-per-capita (Data Source: FAO, 2017)

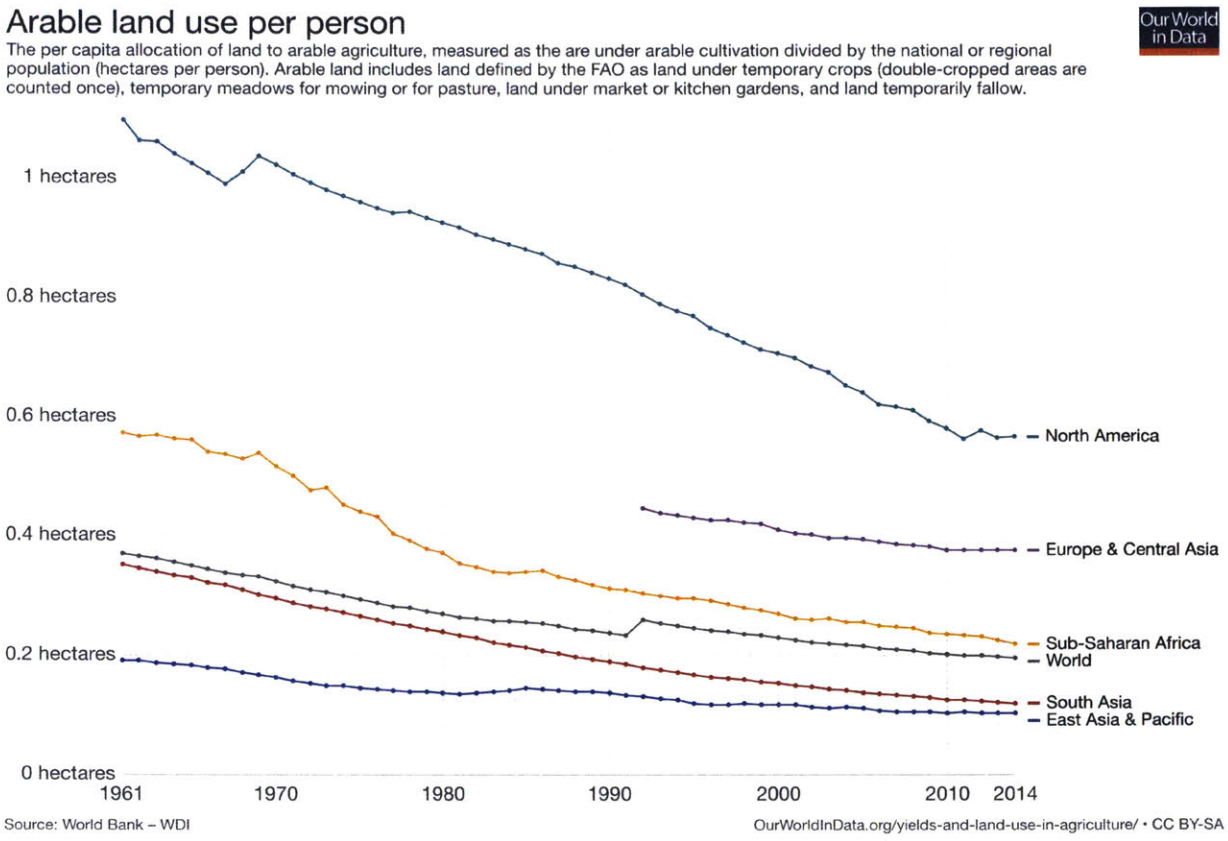


Figure B-5: Arable land use per person

Figure available at https://ourworldindata.org/grapher/arable-land-use-per-person (Data Source: World Bank, 2017)

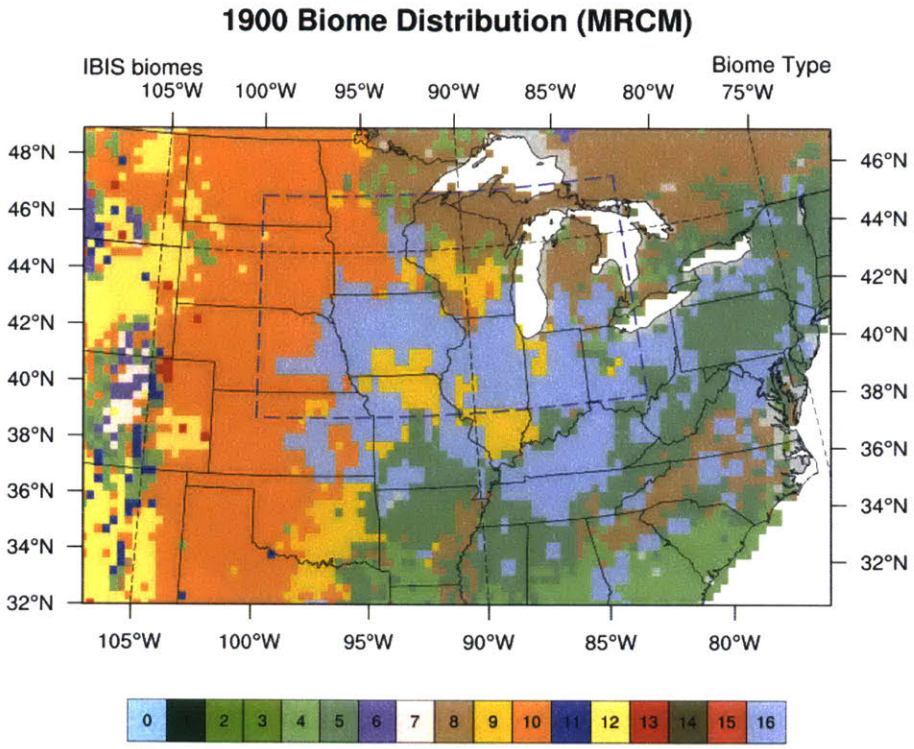


Figure B-6: MRCM Biome Distribution for 1900

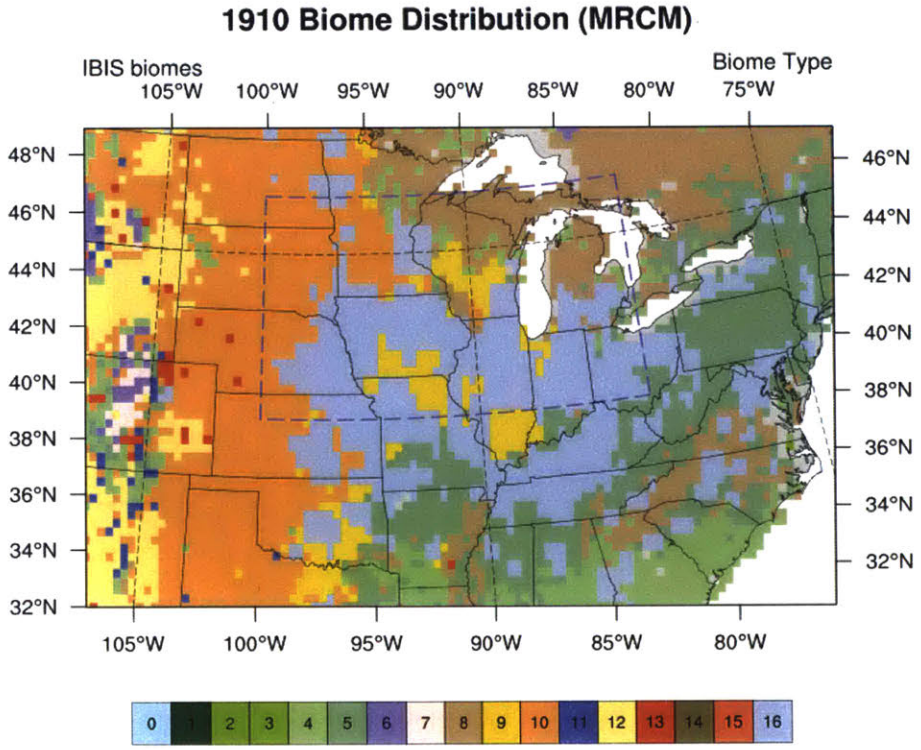


Figure B-7: MRCM Biome Distribution for 1910

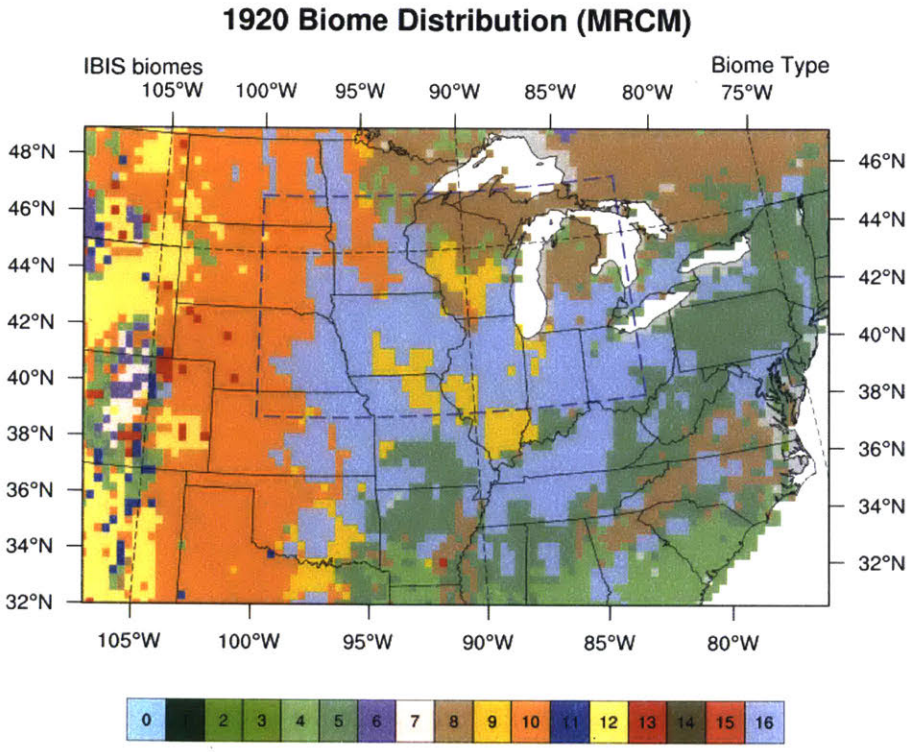


Figure B-8: MRCM Biome Distribution for 1920

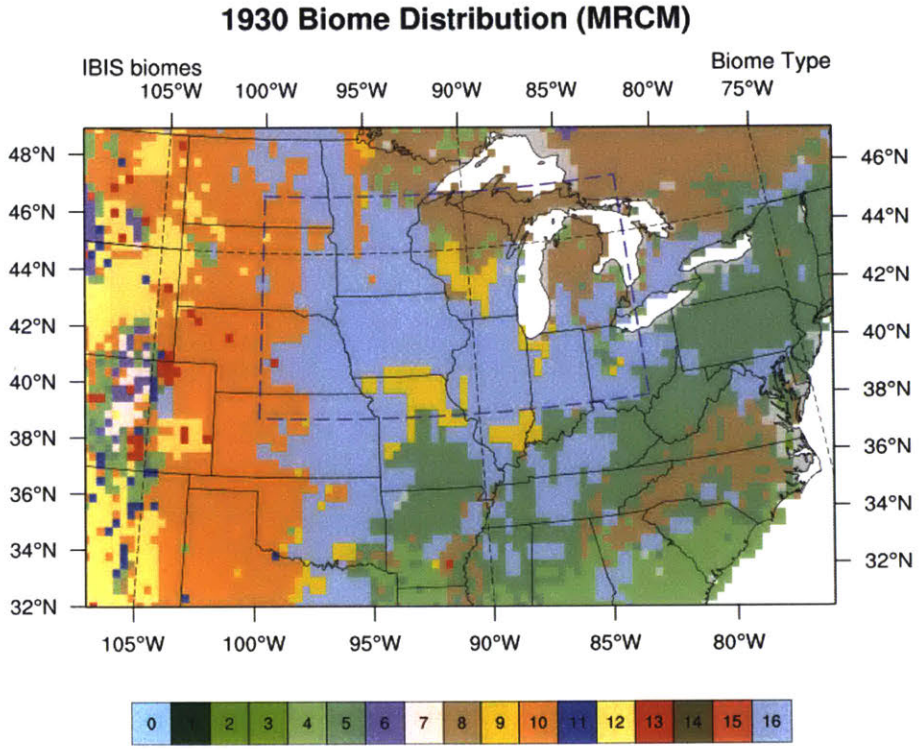


Figure B-9: MRCM Biome Distribution for 1930

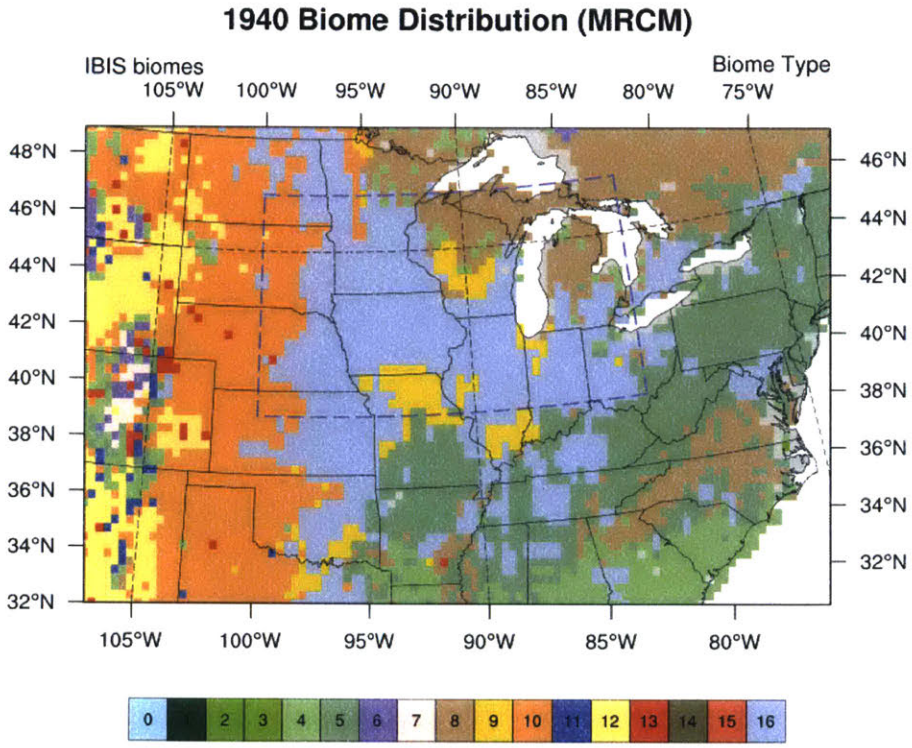


Figure B-10: MRCM Biome Distribution for 1940

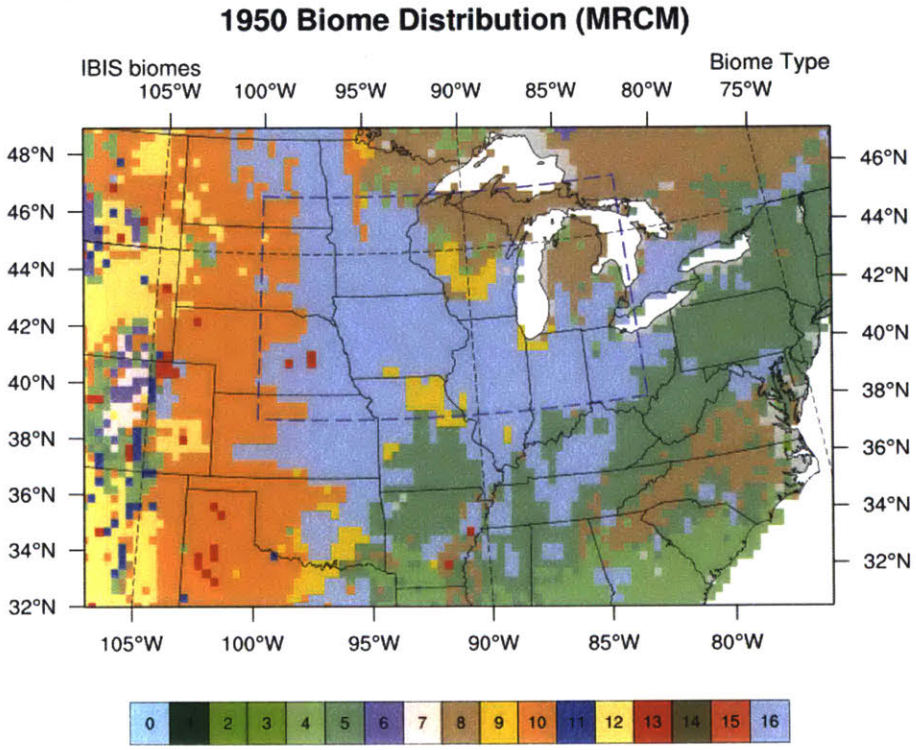


Figure B-11: MRCM Biome Distribution for 1950

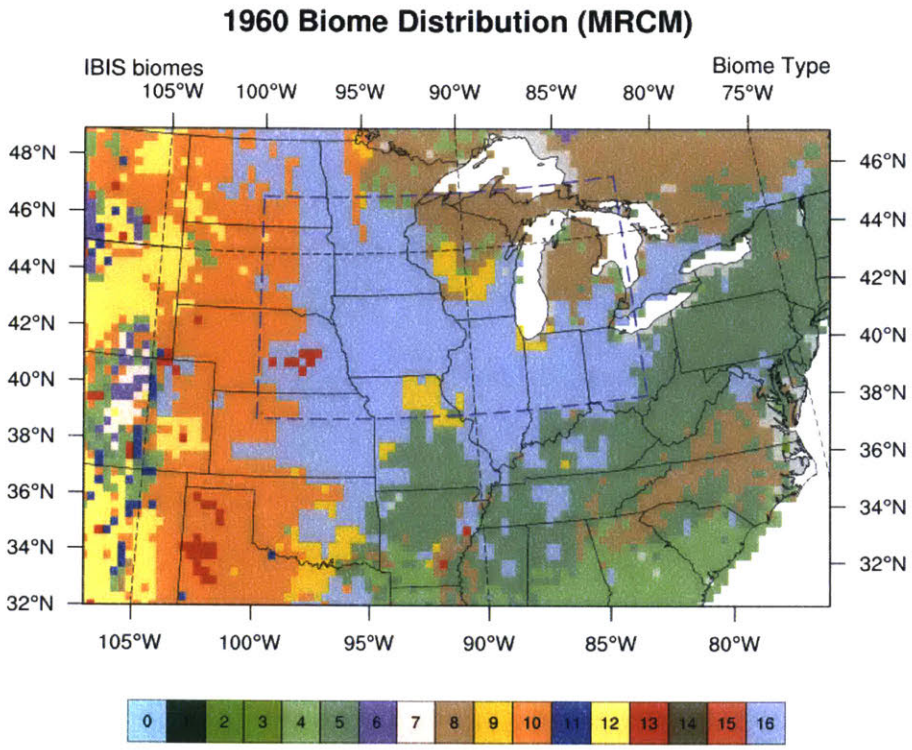


Figure B-12: MRCM Biome Distribution for 1960

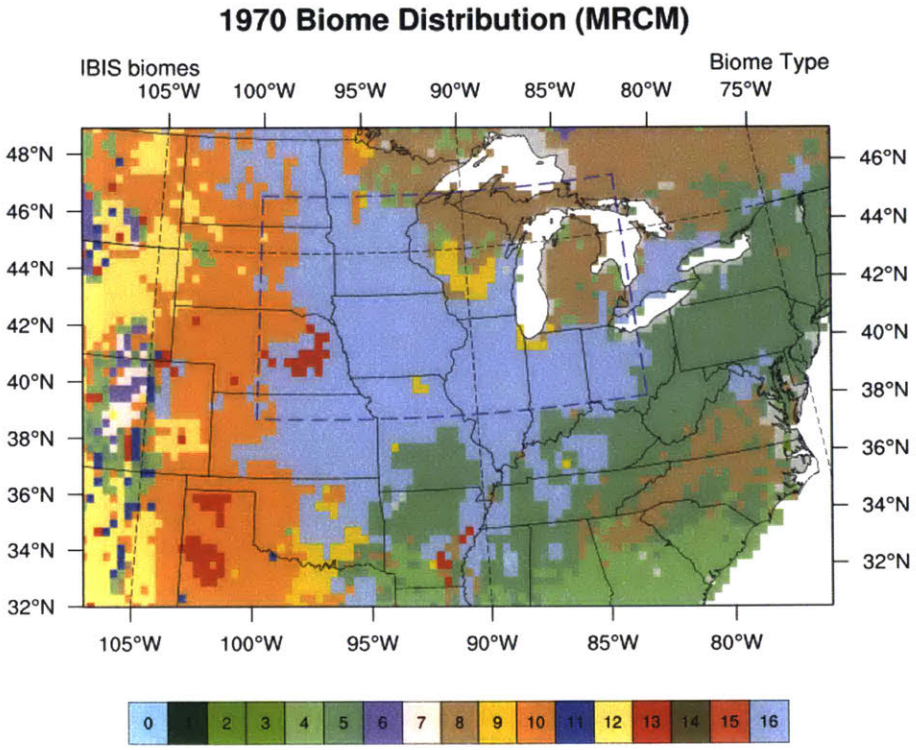


Figure B-13: MRCM Biome Distribution for 1970

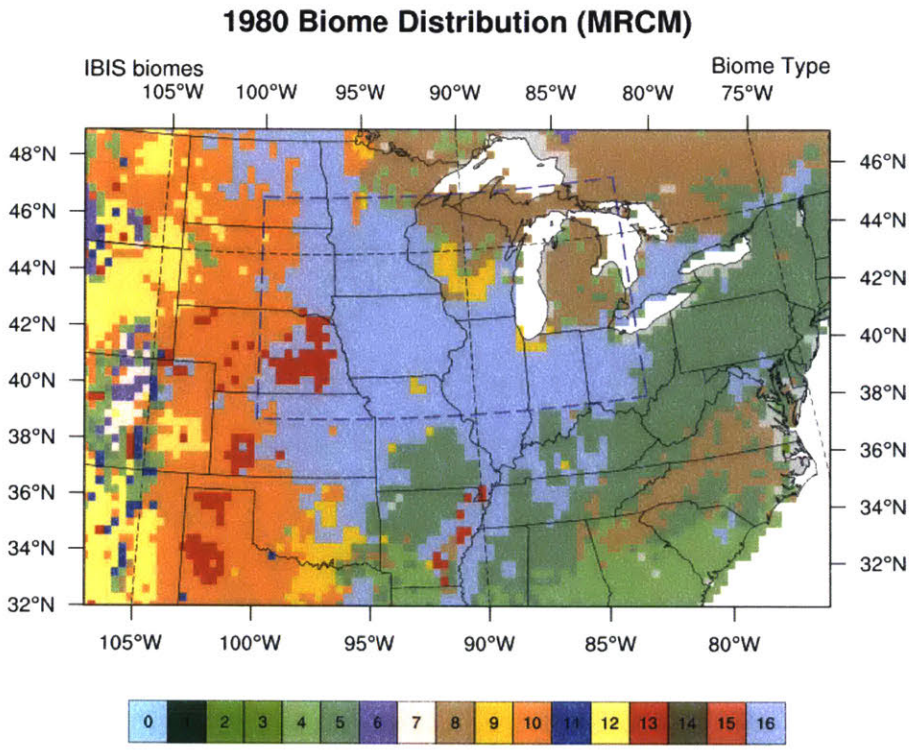


Figure B-14: MRCM Biome Distribution for 1980

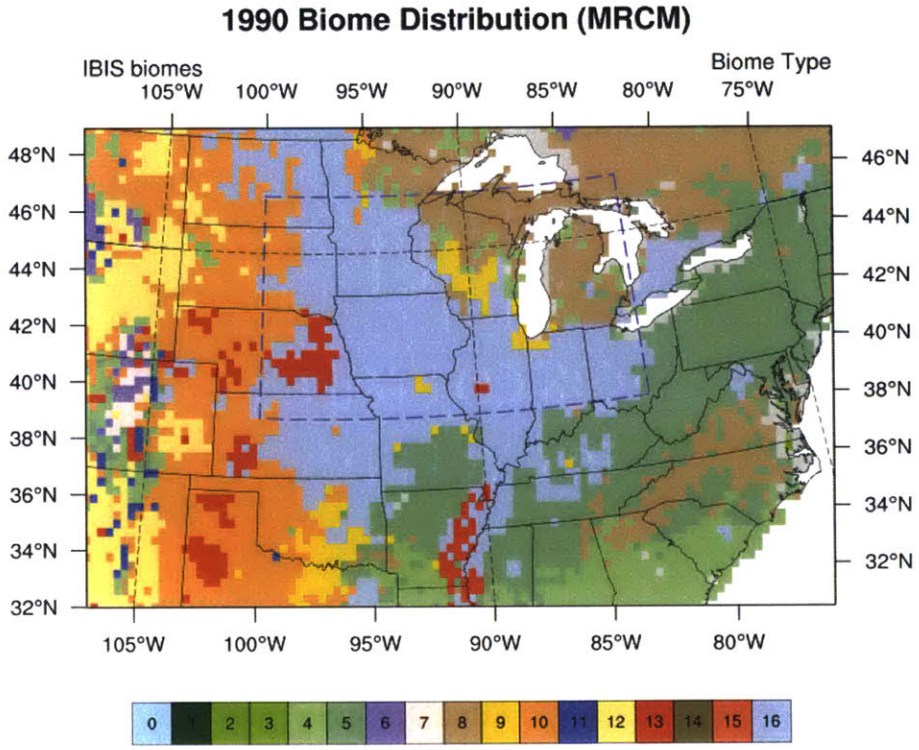


Figure B-15: MRCM Biome Distribution for 1990

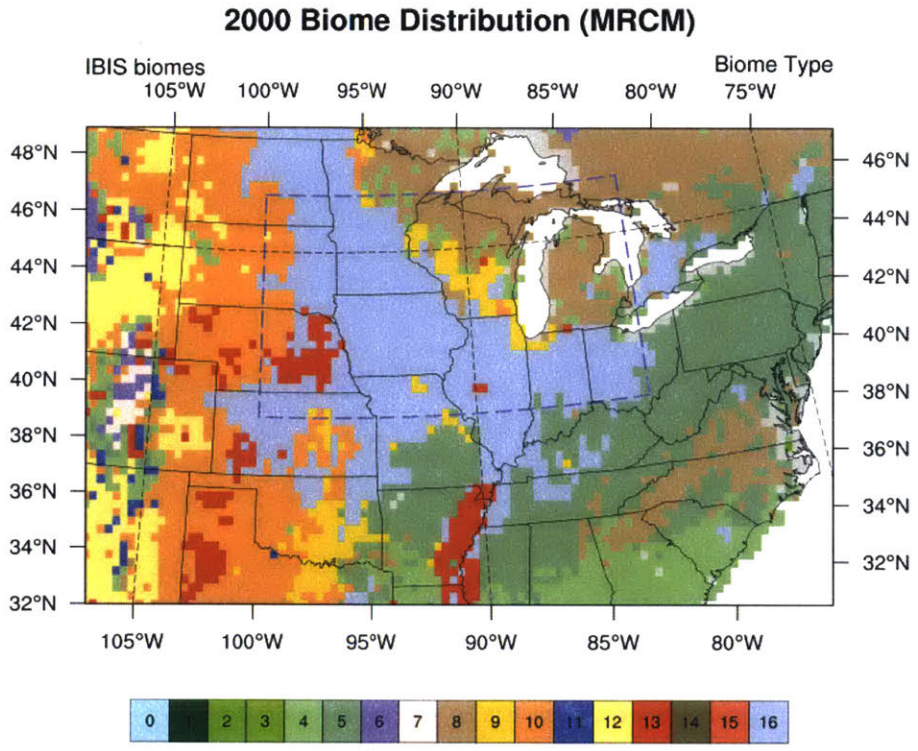


Figure B-16: MRCM Biome Distribution for 2000

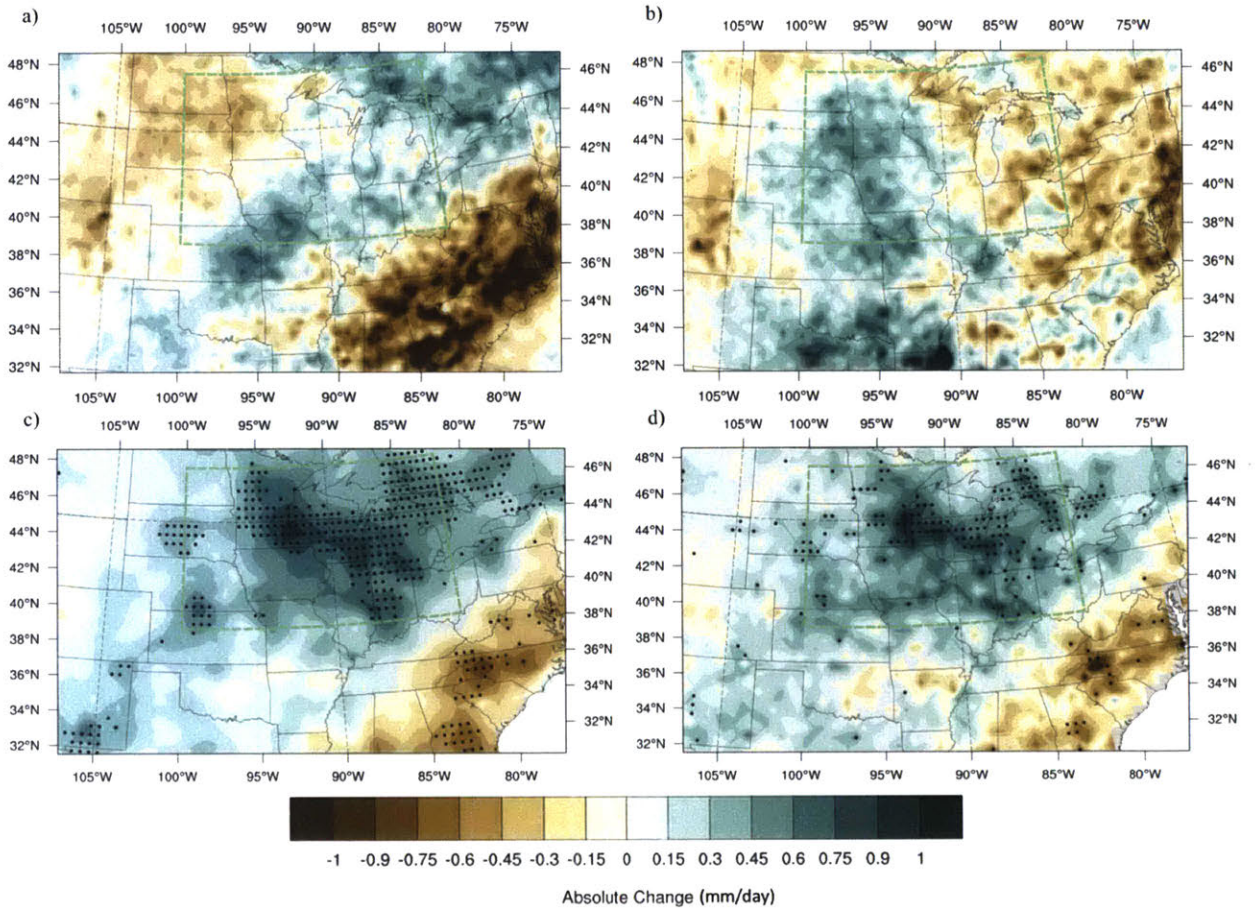


Figure B-17: Absolute average precipitation change (mm/day) from 1920-1949 to 1970-1999 in a) CERA20C VG b) ERA20C VG c) CRUTS4.01 d) UDELv4.01. The dotted green box denotes the ROSC. Black dots on c) and d) indicate grid cells with significant change at the 5% level with a K-S Test.

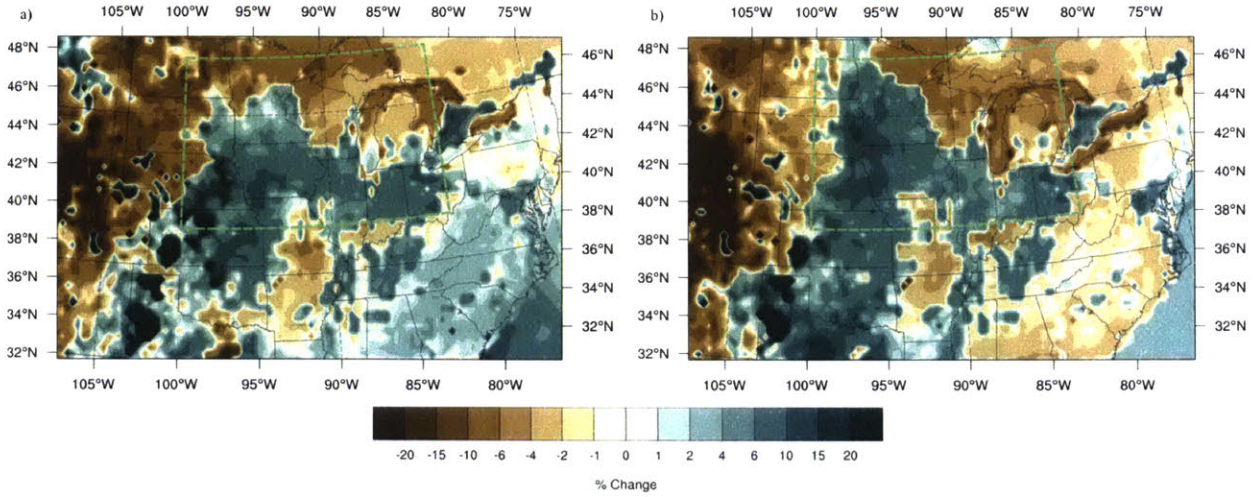


Figure B-18: Average evapotranspiration change (mm/day) from 1920-1949 to 1970-1999 in a) CERA20C VG b) ERA20C VG. The dotted green box denotes the ROSC.

1920-1949 ROSC averages are 2.93 and 2.95 mm/day and the domain averages are 2.74 and 2.82 mm/day respectively. 1970-1999 ROSC averages are 2.96 and 3.01 mm/day and the domain averages are 2.77 and 2.83 mm/day. Therefore, the CERA20C and ERA20C changes from the early to the late period are 0.0333 and 0.0571 mm/day in the ROSC and 0.0281 and 0.0146 mm/day in the domain. None of these changes are significant in the ROSC ($p = 0.5372$ and $p = 0.3420$) nor the domain ($p = 0.3420$ and $p = 0.7600$)

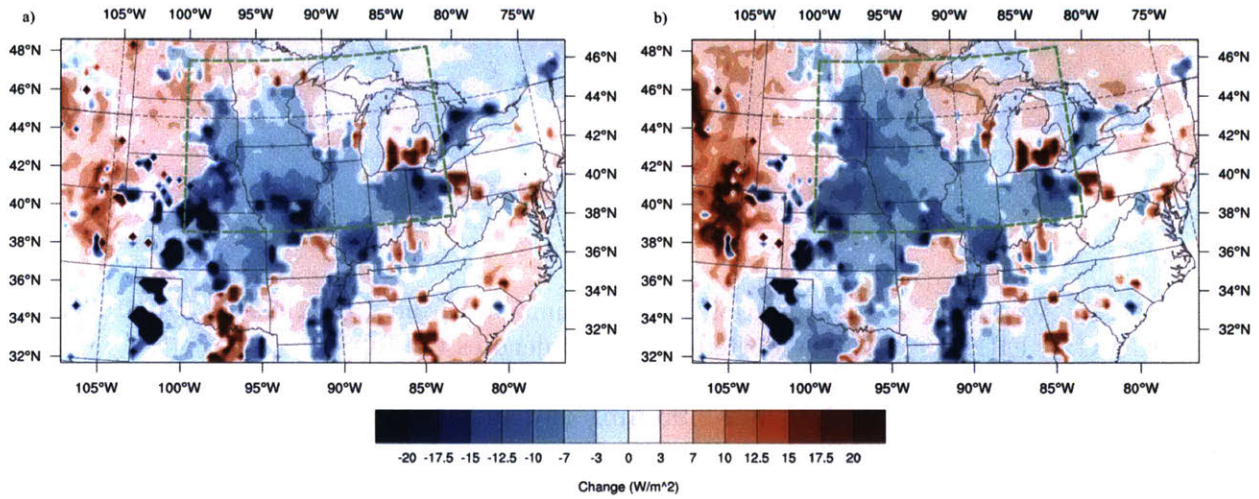


Figure B-19: Average sensible heat change (W/m^2) from 1920-1949 to 1970-1999 in a) CERA20C VG b) ERA20C VG. The dotted green box denotes the ROSC.

1920-1949 ROSC averages are 46.96 and 45.57 W/m^2 and the domain averages are 52.80 and 50.76 W/m^2 respectively. 1970-1999 ROSC averages are 44.05 and 43.11 W/m^2 and the domain averages are 52.11 and 50.67 W/m^2. Therefore, the CERA20C and ERA20C changes from the early to the late period are -2.90 and -2.46 W/m^2 in the ROSC and -0.68 and -0.10 W/m^2 in the domain. These changes are significant in the ROSC ($p = 0.0017$ and $p = 0.0046$) but not in the domain ($p = 0.9260$ and $p = 0.9360$)

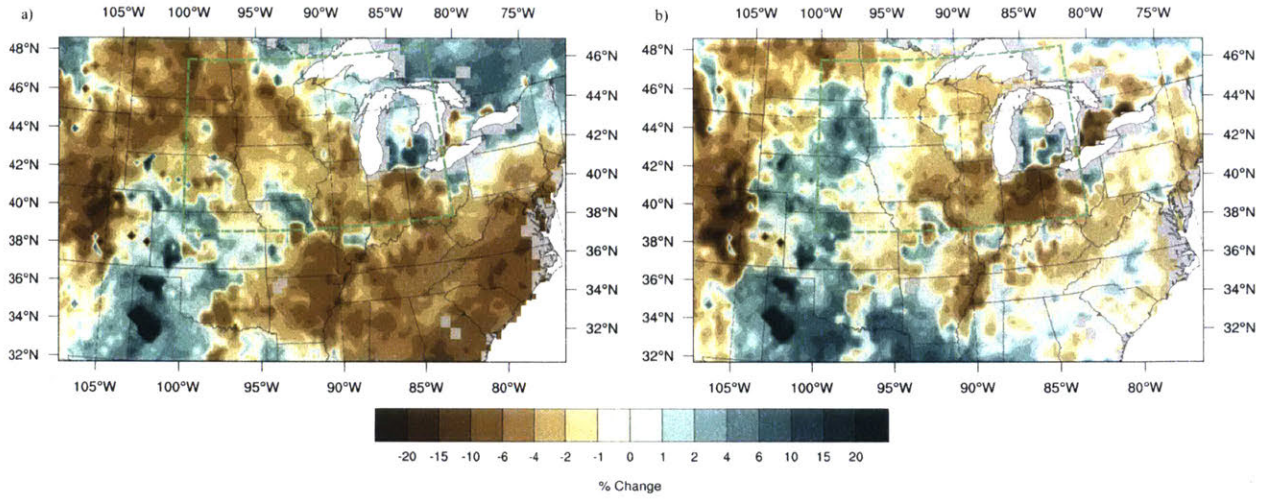


Figure B-20: Average surface soil moisture change (%) from 1920-1949 to 1970-1999 in a) CERA20C VG b) ERA20C VG. The dotted green box denotes the ROSC.

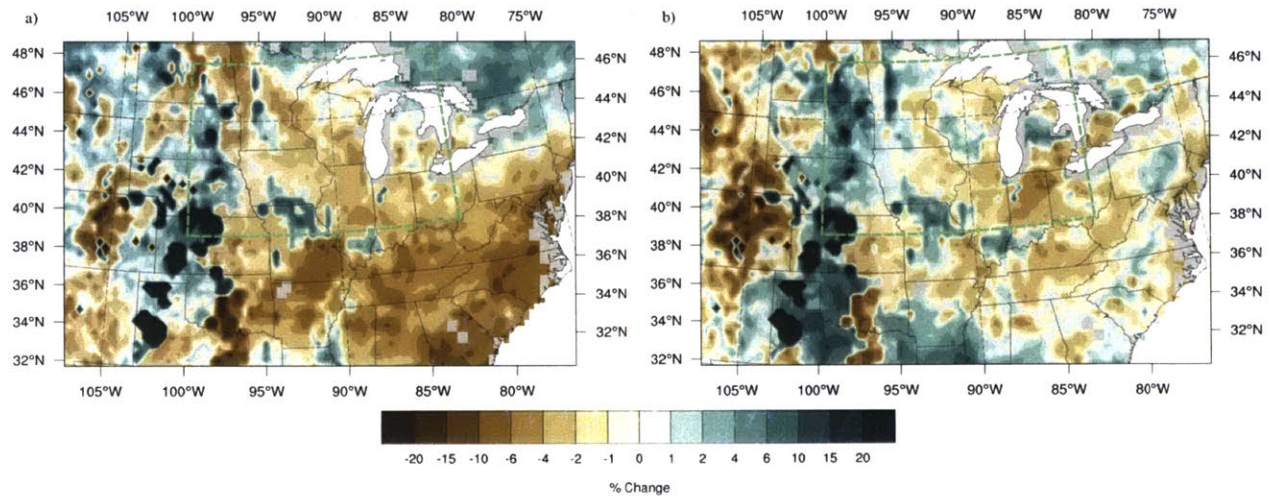
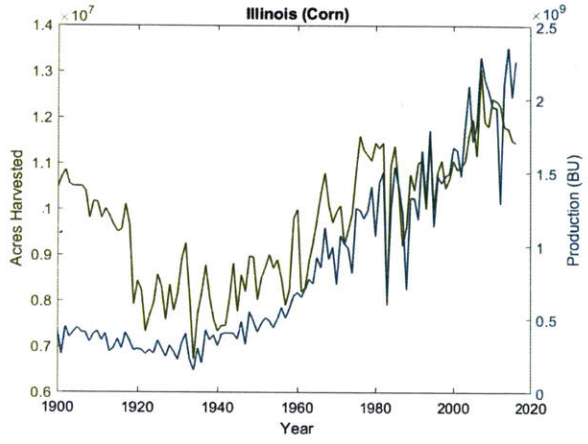


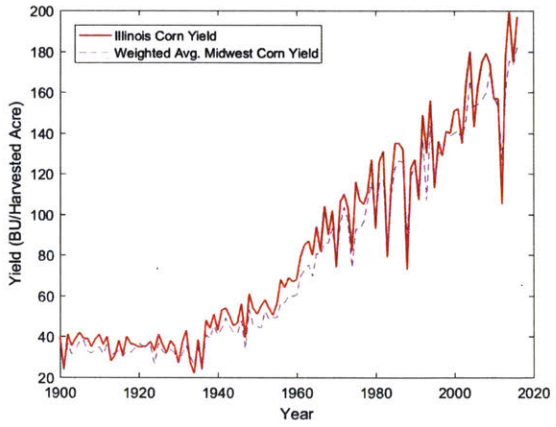
Figure B-21: Average root zone soil moisture change (%) from 1920-1949 to 1970-1999 in a) CERA20C VG b) ERA20C VG. The dotted green box denotes the ROSC.

Appendix C

Corn and Soybean Yield by State

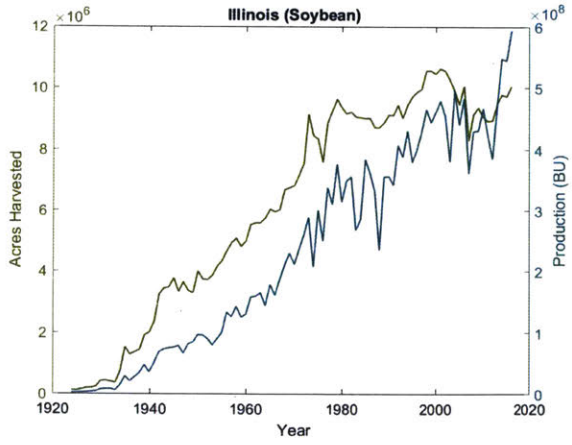


(a) Illinois Corn Acreage and Production

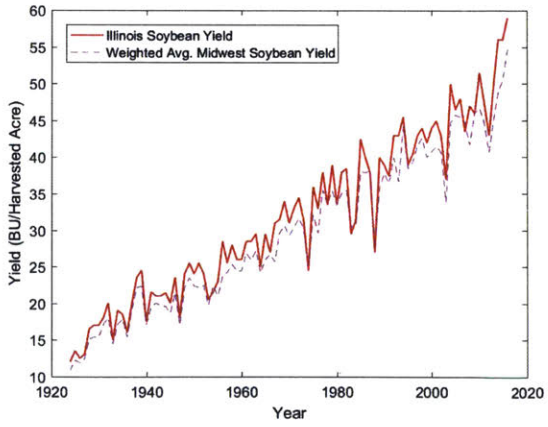


(b) Illinois Corn Yield

Figure C-1: Illinois: Corn

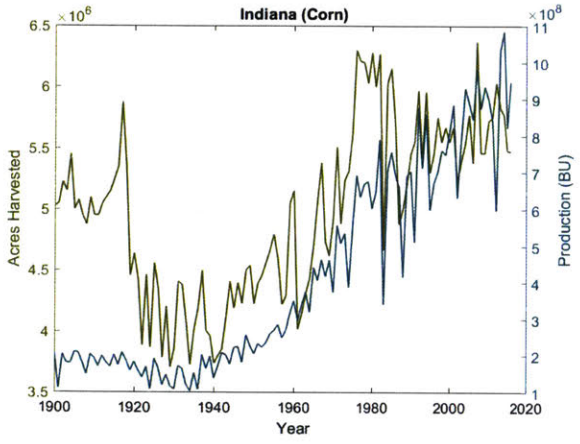


(a) Illinois Soybean Acreage and Production

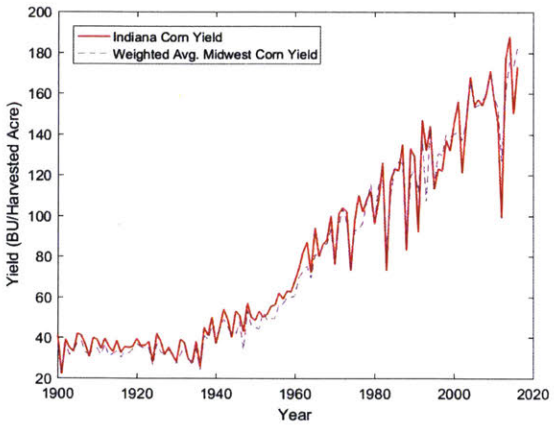


(b) Illinois Soybean Yield

Figure C-2: Illinois: Soybean

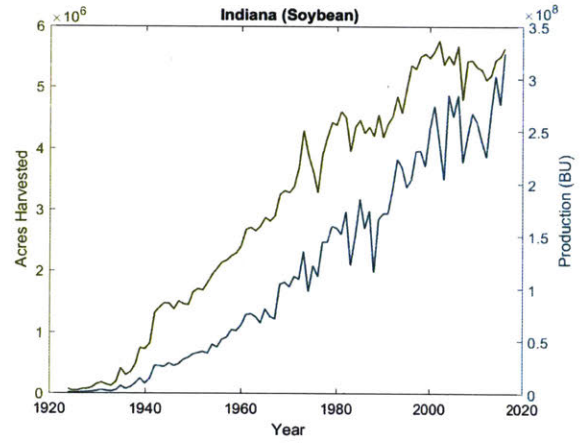


(a) Indiana Corn Acreage and Production

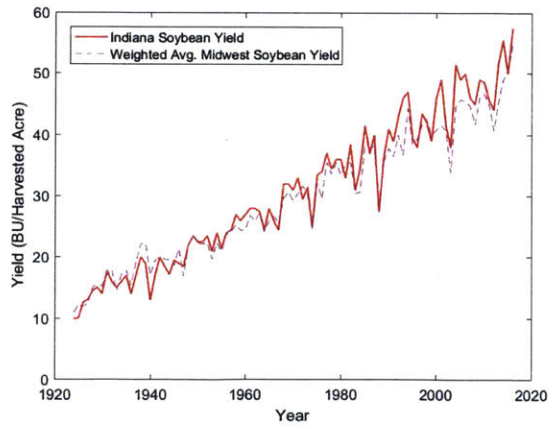


(b) Indiana Corn Yield

Figure C-3: Indiana: Corn

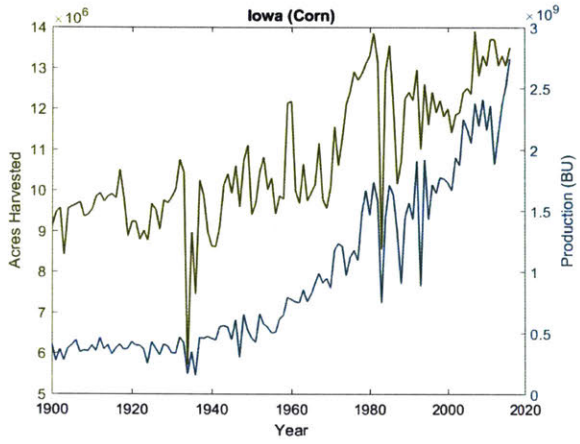


(a) Indiana Soybean Acreage and Production

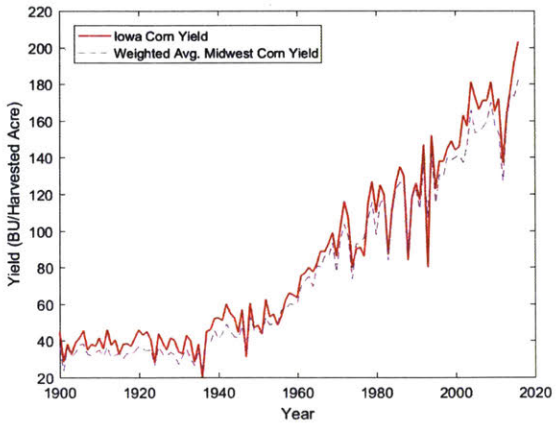


(b) Indiana Soybean Yield

Figure C-4: Indiana: Soybean

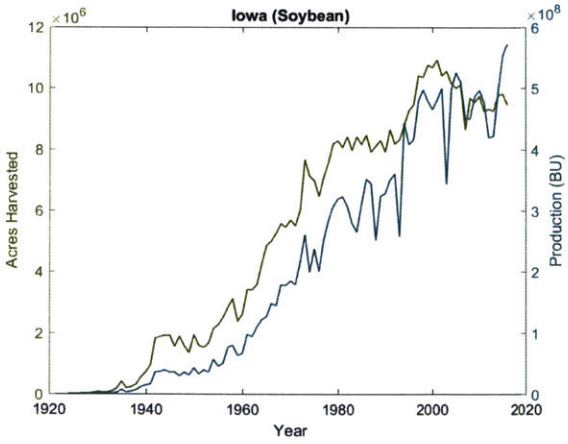


(a) Iowa Corn Acreage and Production

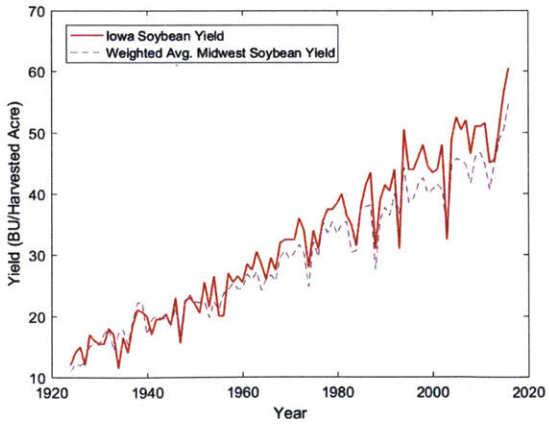


(b) Iowa Corn Yield

Figure C-5: Iowa: Corn

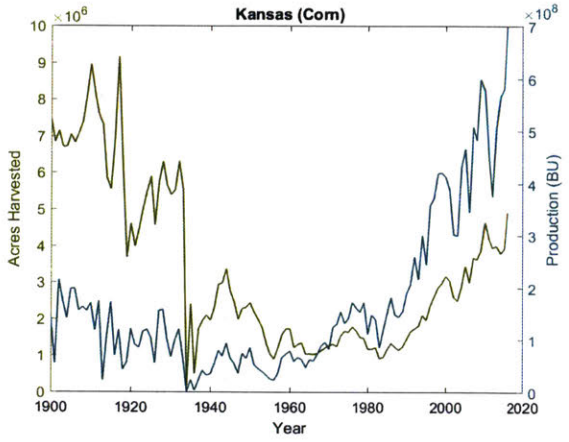


(a) Iowa Soybean Acreage and Production

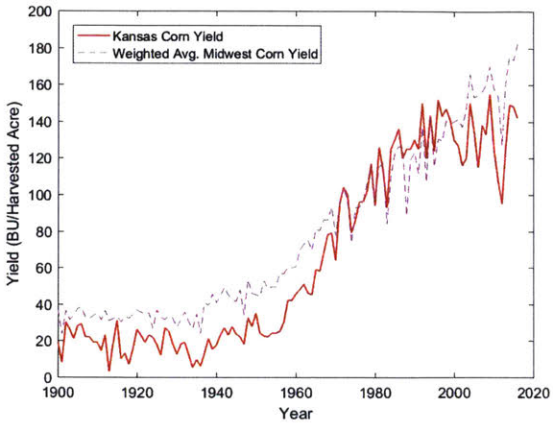


(b) Iowa Soybean Yield

Figure C-6: Iowa: Soybean

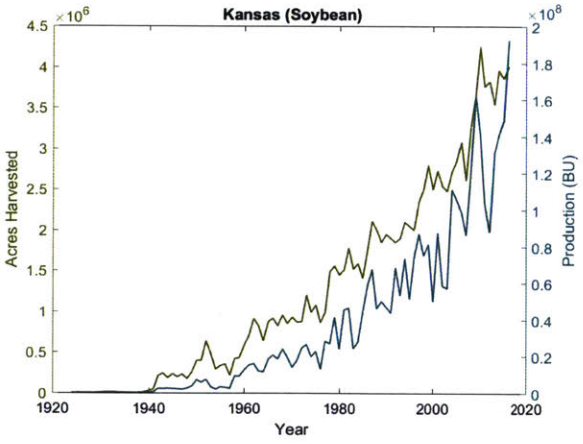


(a) Kansas Corn Acreage and Production

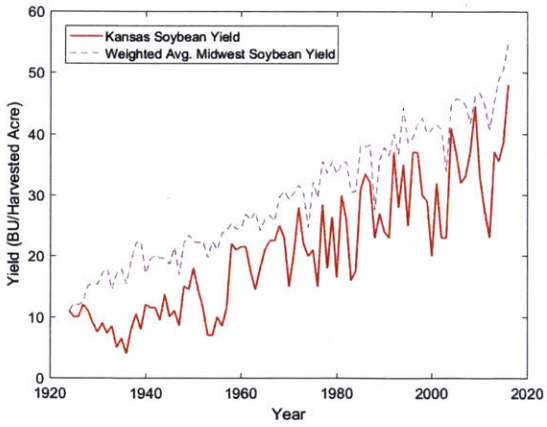


(b) Kansas Corn Yield

Figure C-7: Kansas: Corn

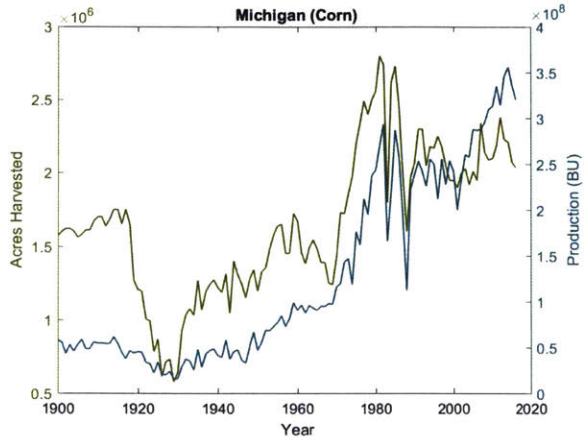


(a) Kansas Soybean Acreage and Production

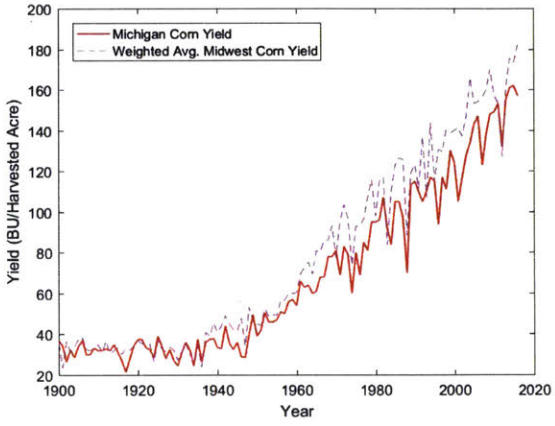


(b) Kansas Soybean Yield

Figure C-8: Kansas: Soybean

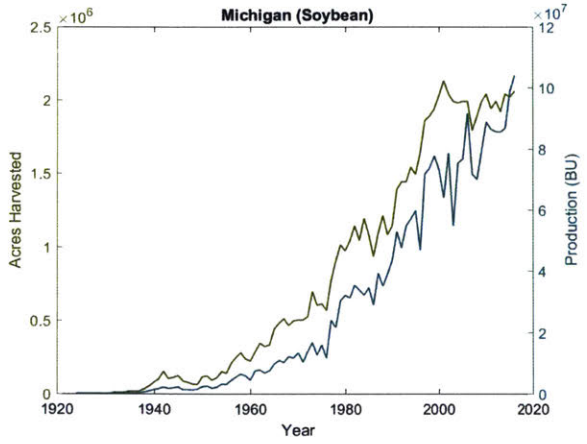


(a) Michigan Corn Acreage and Production

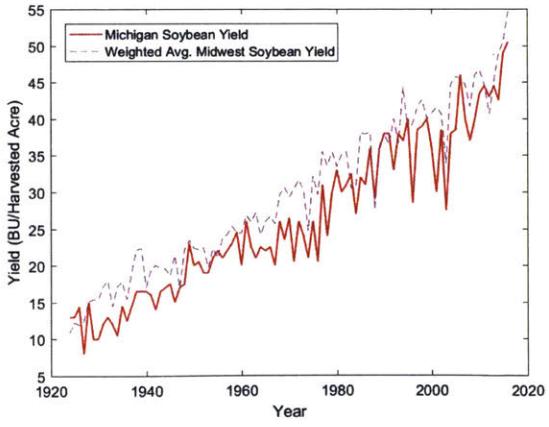


(b) Michigan Corn Yield

Figure C-9: Michigan: Corn

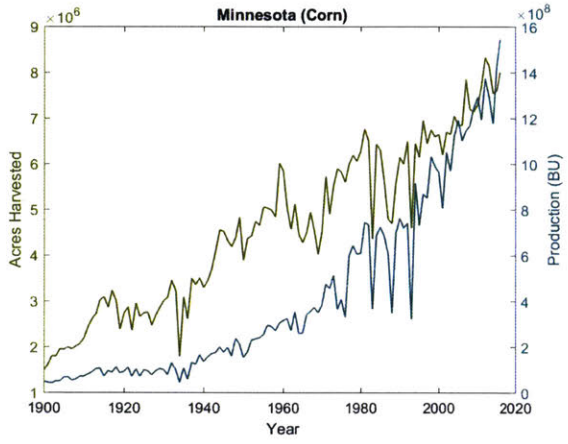


(a) Michigan Soybean Acreage and Production

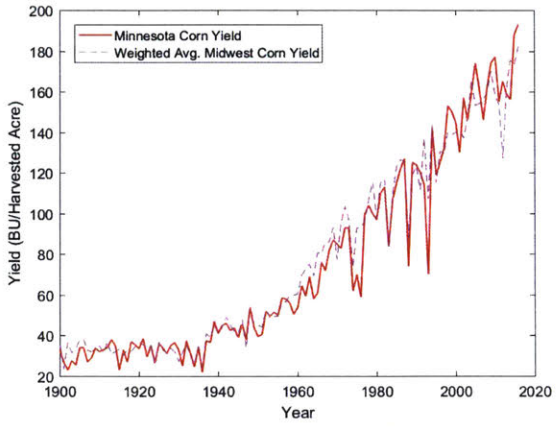


(b) Michigan Soybean Yield

Figure C-10: Michigan: Soybean

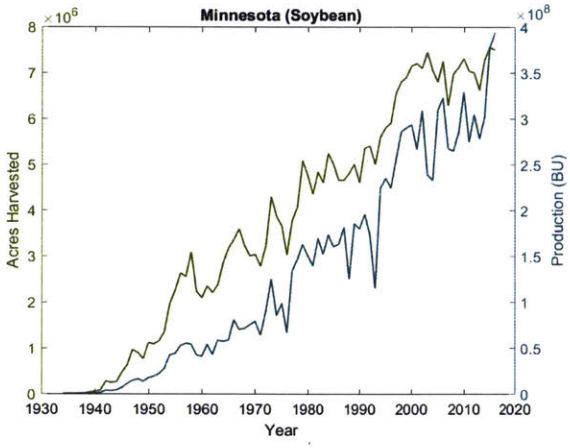


(a) Minnesota Corn Acreage and Production

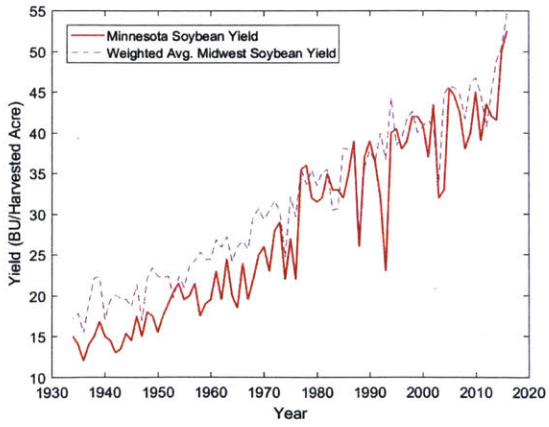


(b) Minnesota Corn Yield

Figure C-11: Minnesota: Corn

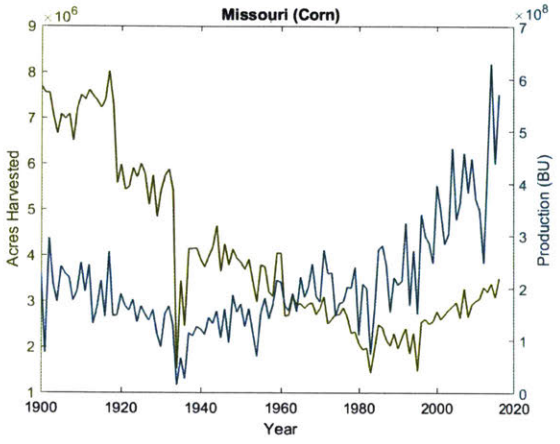


(a) Minnesota Soybean Acreage and Production

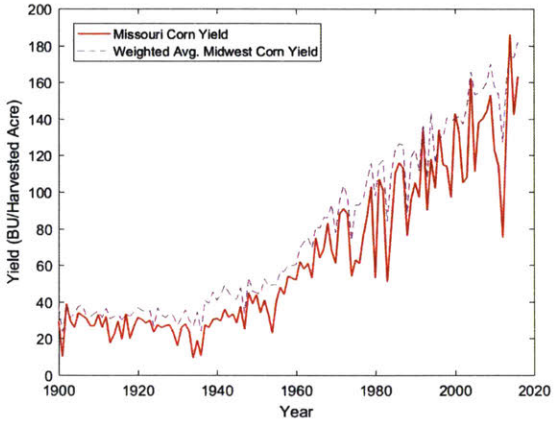


(b) Minnesota Soybean Yield

Figure C-12: Minnesota: Soybean

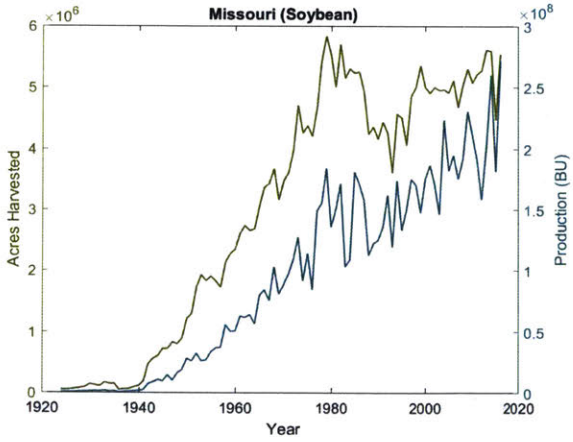


(a) Missouri Corn Acreage and Production

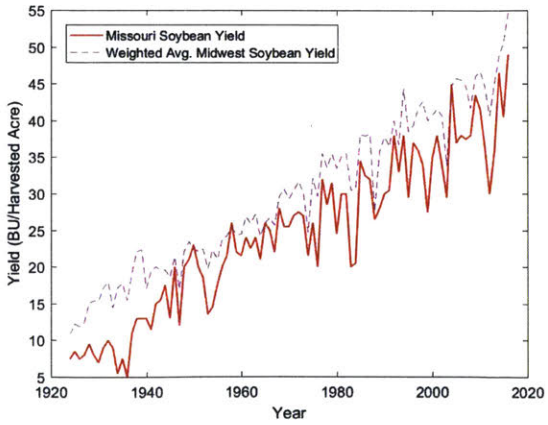


(b) Missouri Corn Yield

Figure C-13: Missouri: Corn

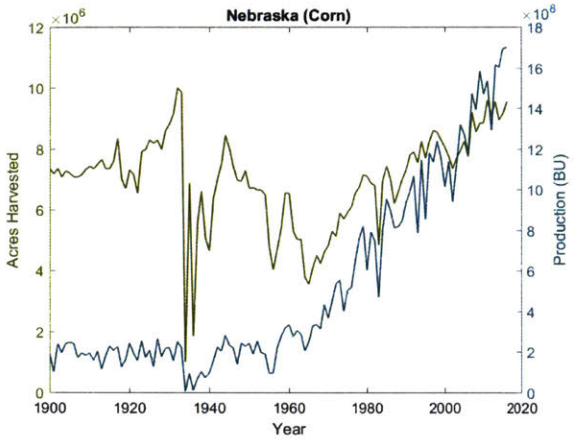


(a) Missouri Soybean Acreage and Production

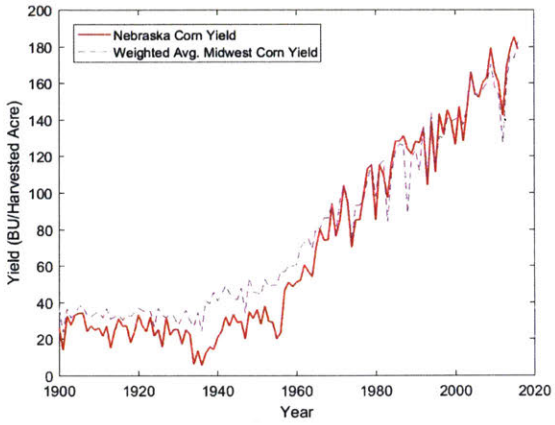


(b) Missouri Soybean Yield

Figure C-14: Missouri: Soybean

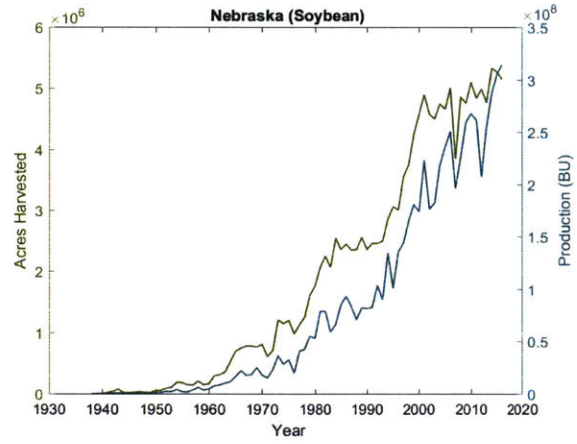


(a) Nebraska Corn Acreage and Production

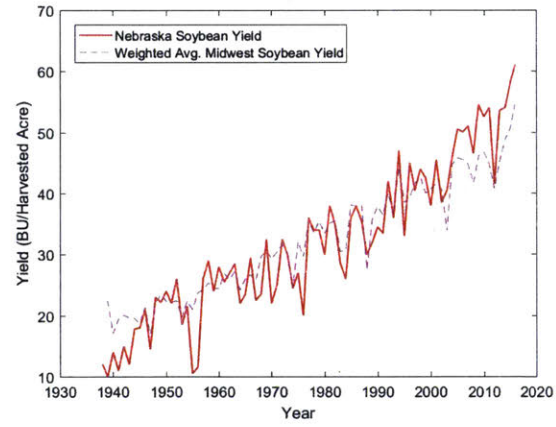


(b) Nebraska Corn Yield

Figure C-15: Nebraska: Corn

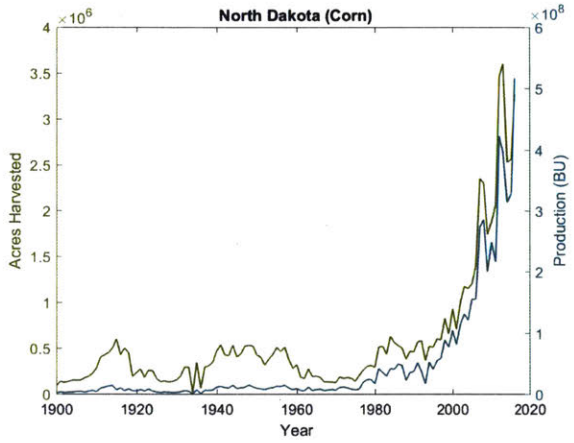


(a) Nebraska Soybean Acreage and Production

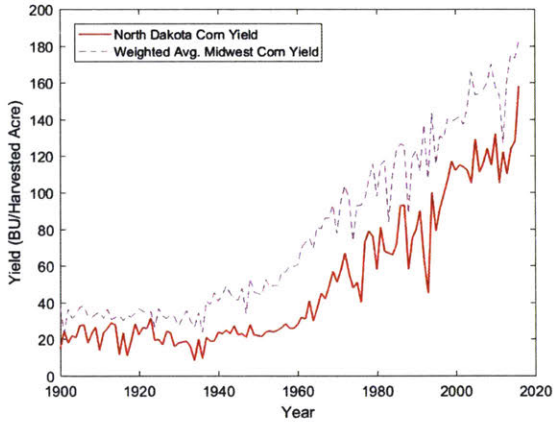


(b) Nebraska Soybean Yield

Figure C-16: Nebraska: Soybean

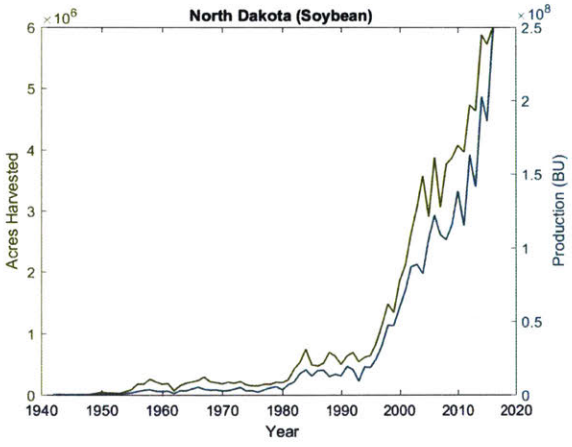


(a) North Dakota Corn Acreage and Production

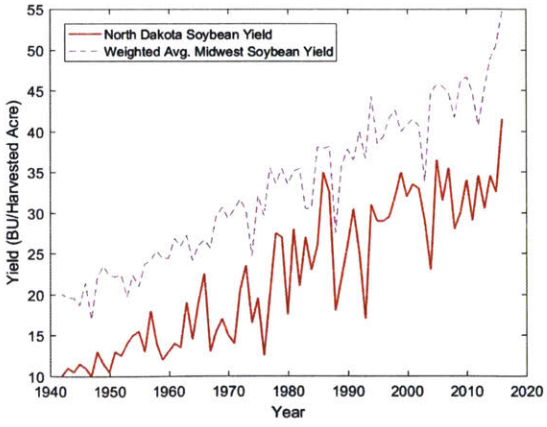


(b) North Dakota Corn Yield

Figure C-17: North Dakota: Corn

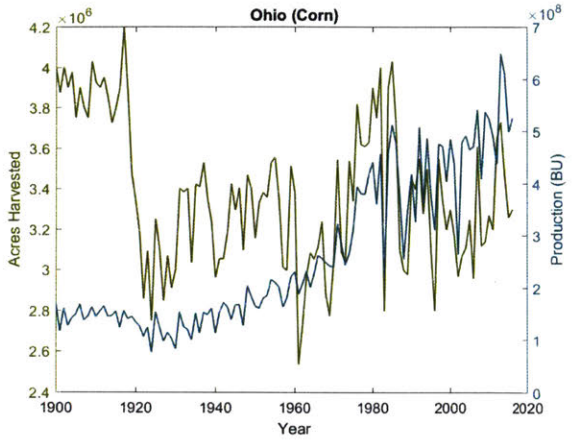


(a) North Dakota Soybean Acreage and Production

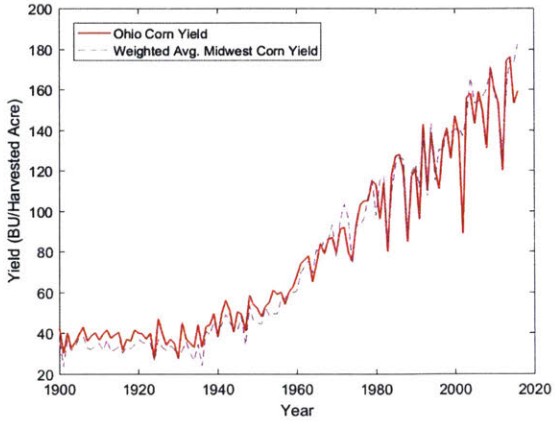


(b) North Dakota Soybean Yield

Figure C-18: North Dakota: Soybean

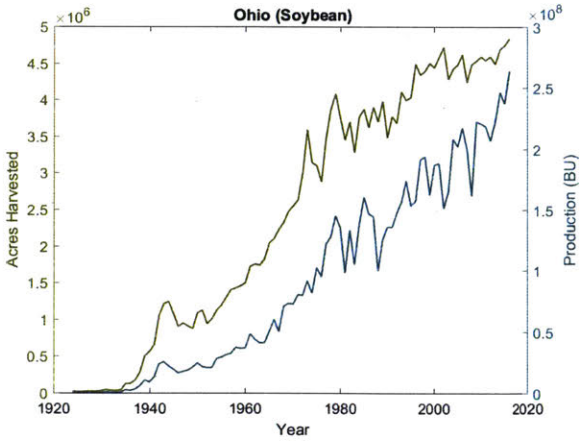


(a) Ohio Corn Acreage and Production

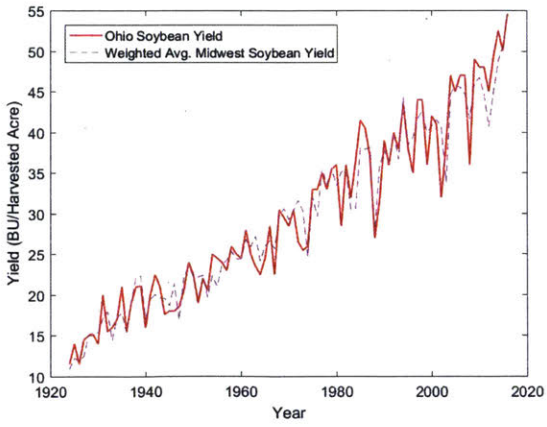


(b) Ohio Corn Yield

Figure C-19: Ohio: Corn

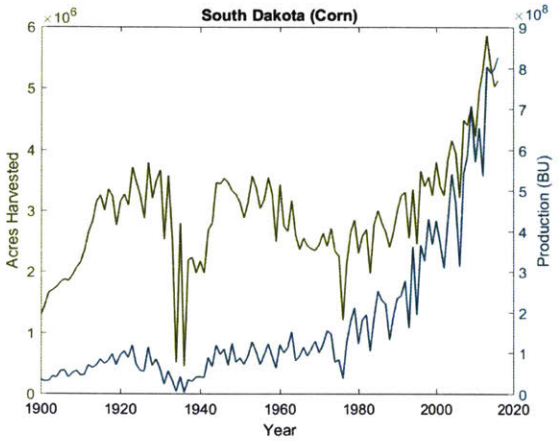


(a) Ohio Soybean Acreage and Production

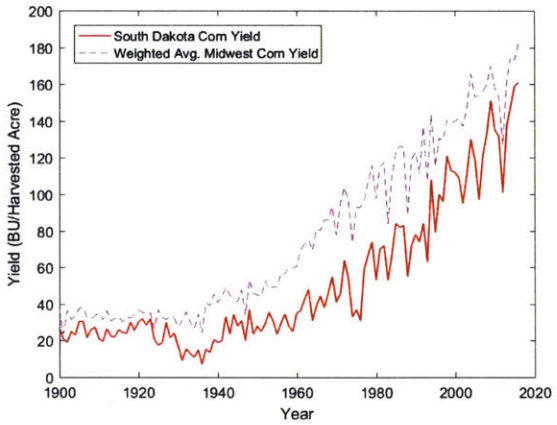


(b) Ohio Soybean Yield

Figure C-20: Ohio: Soybean

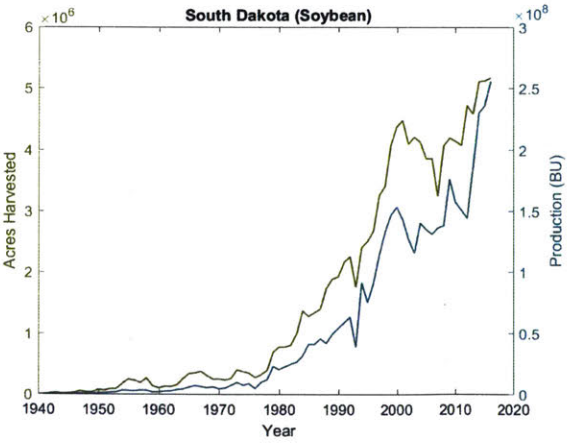


(a) South Dakota Corn Acreage and Production

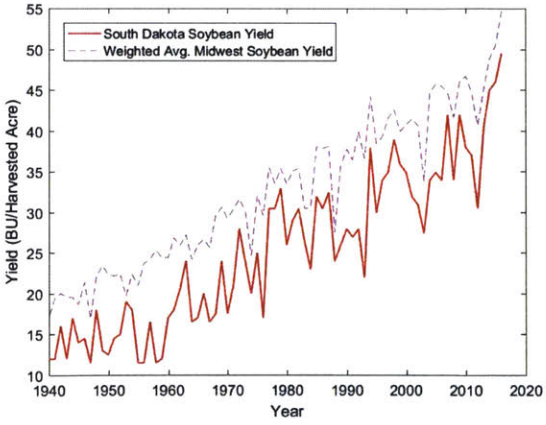


(b) South Dakota Corn Yield

Figure C-21: South Dakota: Corn

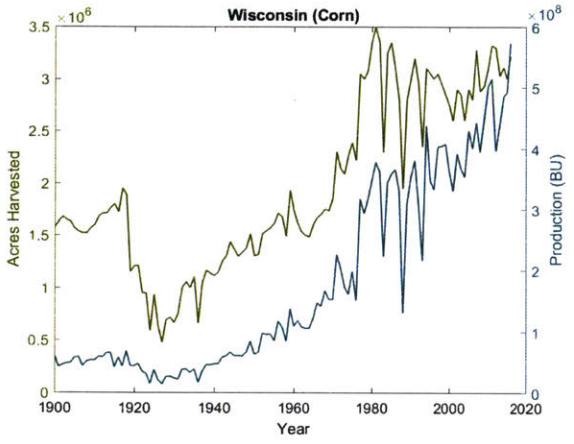


(a) South Dakota Soybean Acreage and Production

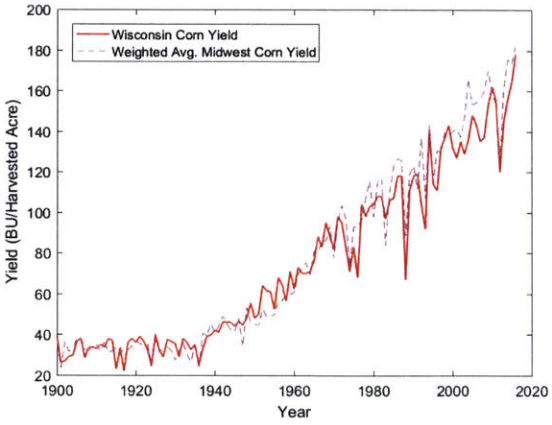


(b) South Dakota Soybean Yield

Figure C-22: South Dakota: Soybean

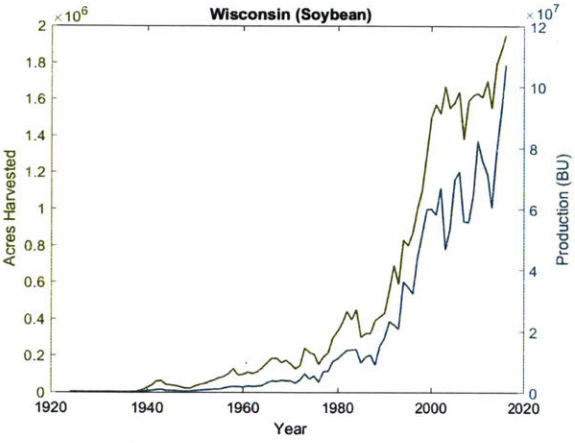


(a) Wisconsin Corn Acreage and Production

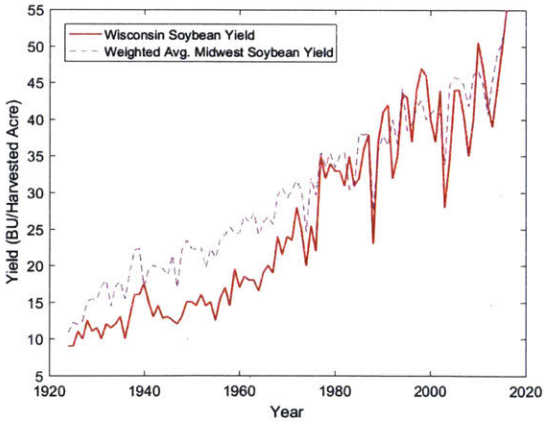


(b) Wisconsin Corn Yield

Figure C-23: Wisconsin: Corn



(a) Wisconsin Soybean Acreage and Production



(b) Wisconsin Soybean Yield

Figure C-24: Wisconsin: Soybean